А. Д. НАСЛЕДОВ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПСИХОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Анализ и интерпретация данных



А. Д. НАСЛЕДОВ

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПСИХОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Анализ и интерпретация данных

Учебное пособие



Санкт-Петербург
2004

ББК 88.36
Н31

Рецензенты:
В.М. Аллахвердов, доктор психологических наук, профессор кафедры общей психологии СПбГУ;
В. М. Буре, кандидат физико-математических наук, доцент факультета прикладной математики — процессов управления СПбГУ

Рекомендовано
Ученым советом факультета психологии СПбГУ
в качестве учебного пособия

Наследов А. Д.
Н31 Математические методы психологического исследования. Анализ и интерпретация данных. Учебное пособие. — СПб.: Речь, 2004. — 392 с.

ISBN 5-9268-0275-7

В данной книге многообразие математико-статистических методов представлено в виде упорядоченной, логически и иерархически взаимосвязанной системы с ориентацией на читателя, не имеющего основательной математической подготовки. Описаны основы применения этих методов, алгоритмы их выбора в зависимости от исследовательской ситуации — от исходных данных и задач исследования. При изложении методов основное внимание уделяется границам их применения, возможным альтернативам, особенностям интерпретации результатов. Применение каждого метода сопровождается простыми примерами и пошаговыми алгоритмами вычислений — как «вручную», так и на компьютере.

Книга адресована студентам психологических и педагогических специальностей, но может быть полезна и широкому кругу исследователей как справочник и руководство по обработке данных.

ББК 88.36

ISBN 5-9268-0275-7

КРАТКОЕ СОДЕРЖАНИЕ

Часть I
ОСНОВЫ ИЗМЕРЕНИЯ И КОЛИЧЕСТВЕННОГО ОПИСАНИЯ ДАННЫХ

Часть II
МЕТОДЫ СТАТИСТИЧЕСКОГО ВЫВОДА: ПРОВЕРКА ГИПОТЕЗ

Часть III
МНОГОМЕРНЫЕ МЕТОДЫ И МОДЕЛИ

ОГЛАВЛЕНИЕ

Часть III
МНОГОМЕРНЫЕ МЕТОДЫ И МОДЕЛИ

Приложения
ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

ПРЕДИСЛОВИЕ

Среди других наук психологию отличает интригующее разнообразие точек зрения на перспективы ее развития. Оставляю другим увлекательные возможности выражать свои сомнения и жонглировать словами на сей счет. При написании этой книги я следовал вполне традиционному убеждению: психология в любых ее приложениях — и практических, и теоретических, может развиваться только на основе количественных исследований, связывающих теорию и практику с фактами.

Исследование в любой области, в том числе и в психологии, предполагает получение результатов — обычно в виде чисел. Однако просто собрать данные недостаточно. Даже объективно и корректно собранные данные ничего не говорят. Исследователю необходимо умение организовать их, обработать и проинтерпретировать, что невозможно без применения математических методов. Конечно, можно сослаться на наличие современных компьютерных программ, применение которых сейчас становится нормой для исследователя. Но любая программа обработки данных переводит один набор чисел в другой набор чисел. При этом предлагается богатый набор способов такого преобразования, замечательным образом расширяющий возможности анализа данных. И для использования этих возможностей психолог должен уметь: а) организовать исследование так, чтобы его результаты были доступны обработке в соответствии с проблемами исследования; б) правильно выбрать метод обработки; в) содержательно интерпретировать результаты обработки. Эти умения не заменят ни компьютерная программа, ни «живой» математик — ее создатель[1]. Таким образом, применение математики как общенаучного метода, наряду с экспериментом, неизбежно приобретает в психологии свои особенности, связанные со спецификой предмета[2]. Неотъемлемой частью подготовки полноценного специалиста-психолога является изучение не только экспериментальной психологии, но и математических методов психологического исследования.

Стиль книги выбран с учетом того, что математические методы обычно вызывают большие трудности при изучении (и преподавании). Студенты (и не только они) часто сомневаются в необходимости изучения математических методов и испытывают страх при мысли о неизбежной перспективе их применения. Поэтому при изложении материала основное внимание уделяется практическим проблемам выбора метода и особенностям интерпретации

[1] Выход есть: заведите себе «ручного» математика! Но учтите: а) на воле такие не водятся, и приручать его вам придется самостоятельно; б) по мере приручения все менее понятно, кто кого приручает...

[2] «Математическая психология» И. Гербарта появилась десятилетиями раньше, чем «экспериментальная психология» В. Вундта.

получаемых результатов. При этом я не стремлюсь к абсолютной математической строгости и доказательности положений, свойственным математическим изданиям. Необходимые для понимания математические основы даются скорее на интуитивном и неформальном уровне — без детального изложения математического обоснования и выводов формул, которые могут вызвать негативные переживания читателя, не обладающего основательной математической подготовкой. Введение математических терминов сопровождается простыми примерами, а теоретические и математические объяснения даются на элементарном уровне. Основные термины в тексте выделены.

Жанр книги по первоначальному замыслу — учебное пособие для студентов факультета психологии. Но в процессе работы над книгой источником идей являлась не только практика преподавания, но и опыт участия в многочисленных исследованиях в роли руководителя или консультанта. В итоге появились основания надеяться, что книга станет не только *учебником* для студентов, но будет полезна для широкого круга исследователей — как *справочник* и *практическое руководство* по анализу и интерпретации данных. Справочному назначению книги способствует предметный указатель и англо-русский терминологический словарь, а практическое руководство воплощено в пошаговых инструкциях по применению каждого из методов.

Назначение книги — формирование умений самостоятельно анализировать и, главное, интерпретировать эмпирические данные — результаты исследований. Как пишет Г. В. Суходольский: «В психологии следует различать и уметь выполнять четыре вида интерпретаций: психолого-психологические, психолого-математические, математико-математические и (обратные) математико-психологические»[1]. *Психолого-математическая* интерпретация заключается в математической идентификации исследовательской ситуации, которая сводится к выбору методов анализа данных. Решению этой проблемы способствуют предлагаемые классификации математических методов (статистического вывода и многомерных методов). Их особенность заключается в том, что исходным основанием для выбора адекватного метода (и его альтернатив) является специфика исследовательской ситуации, а не математическая специфика метода. Изложение каждого метода сопровождается обсуждением границ его применения и возможных альтернатив. *Математико-математической интерпретации* соответствуют вычисления: переход при помощи выбранного метода от длинной исходной последовательности чисел к более короткому их набору — результатам обработки. Вычислительному аспекту в книге уделено особое внимание — чтобы читатель смог самостоятельно произвести все необходимые расчеты. При описании каждого метода в общих чертах объясняются его математические основы, демонстрируются примеры применения и, главное, предлагается *пошаговый алгоритм вычислений*: для обработки «вручную» (если это возможно) и с помощью компьютера. *Математико-психологическая интерпретация* (допустимая и возможная содержательная интерпретация числовых результатов) находится в центре внимания при рассмотрении любого метода. Примерно половина содержания книги — это обсуждение вопроса: «Что можно сказать по поводу вычисленных показателей?». Все аспекты интерпретации иллюстрированы

---

[1] *Суходольский Г. В.* Математические методы психологии. СПб., 2003. С. 242.

многочисленными примерами. При этом наряду с реальными случаями чаще используются вымышленные данные и примеры — для наглядности и для того, чтобы избавиться от множества второстепенных деталей, неизбежно сопровождающих реальные исследования.

Структура книги соответствует стремлению представить множество математических методов в виде упорядоченной, логически и иерархически взаимосвязанной системы. *Во вступлении* дано общее описание этой системы и ее частей (модели измерения, описания и статистического вывода). Основной материал книги изложен в трех частях. *В первой части* даны элементарные основы применения математических методов. Ее назначение — подготовка читателя к восприятию основного материала книги. Этому способствуют задачи и упражнения в конце глав. *Вторая часть* включает в себя детальное описание основных методов статистического вывода. Их изложение предваряется классификацией, которая позволяет выбрать метод в зависимости от исследовательской ситуации — от исходных данных и задач исследования. При изложении каждого метода особое внимание уделяется границам его применения, возможным альтернативам, технике вычислений («вручную» и на компьютере), особенностям интерпретации результатов. *Третья часть* содержит описание самых распространенных многомерных методов. Применение этих методов возможно только с использованием специальных компьютерных программ. Поэтому их математические основы и порядок вычислений даются лишь в самых общих чертах, а основное внимание уделяется назначению, содержательной интерпретации результатов и, конечно, компьютерной обработке.

Рекомендации читателю зависят от его подготовленности и намерений.

Абитуриенту. Прочитайте «Методологическое введение: психология и математика». Вы все еще желаете поступать на факультет психологии? Если да, то положите эту книгу в доступное место: вам все равно не избежать ее внимательного и неоднократного чтения, будучи студентом. Если нет — найдите 32 забавные картинки в книжке и подумайте, что бы они могли значить.

Студенту. Смело приступайте к чтению с самого начала. Ничего не бойтесь: математики в общеобразовательном ее понимании в этой книги совсем мало. Внимательно читайте примеры и придумывайте свои. Выполняйте задания и упражнения. Приступая к чтению второй и третьей частей: обратитесь к преподавателю, ведущему практические занятия, чтобы он снабдил вас данными для обработки («вручную» и на компьютере). Самостоятельно примените к этим данным каждый из методов. Затем — придумывайте гипотезы, организуйте исследования, обрабатывайте свои данные и интерпретируйте результаты: это интересно! А если по ходу дела возникнут вопросы — вам поможет предметный указатель в конце книги.

Исследователю. Освежите в памяти основы: пролистайте первую часть (обратите внимание на главу 3 — о табличном представлении исходных данных). Внимательно изучите классификацию методов статистического вывода (глава 8) и классификацию многомерных методов (глава 14). Если исследование (сбор данных) только в перспективе, то планируйте его так, чтобы исходные данные позволяли проверить гипотезы вашего исследования. Если данные уже собраны, то с помощью классификаций выберите подходящий метод и приступайте к его изучению: прочитайте соответствующую главу, внимательно

изучите требования к исходным данным, ограничения и пример с пошаговым применением метода. Введите данные в компьютер и убедитесь, что их вид допускает применение метода. Найдите в конце главы с описанием метода раздел «Обработка на компьютере» и следуйте инструкциям. Результаты интерпретируйте по аналогии с примерами. Полезно воспользоваться альтернативными методами и сравнить полученные с их помощью результаты. Если встречаются непонятные термины — обращайтесь к предметному указателю.

Научному руководителю. Сначала просто пролистайте книгу, обратив внимание на главы 8 и 14: вот ведь, оказывается, как велико многообразие доступных способов организации исследования! Затем отдайте эту книгу своему подопечному и помогите ему сформулировать исследовательские гипотезы, которые действительно представляют интерес и доступны проверке. Вам эта книга пригодится потом, когда вы будете оценивать реалистичность исследовательского проекта и далее, при проверке адекватности методов обработки данных и корректности соответствующих выводов.

Убежденному стороннику гуманитарного подхода. Прочитайте «Методологическое вступление...», просмотрите по 1–2 главы из начала каждой части. Я очень надеюсь, что вы поймете: вашему подходу нисколько не противоречит применение количественных методов, которые действительно могут повысить качество и убедительность результатов «качественных» исследований!

Чего нет в этой книге. В этой книге вы не найдете методов анализа результатов «мысленных экспериментов» или «качественных исследований», древнекитайской классификации животных и прочих забавных вещей. А если серьезно, то в книге специально почти не обсуждается начальный этап организации и планирования исследования: это предмет экспериментальной психологии, который соотносится с психолого-психологической интерпретацией (по Г. В. Суходольскому). Имеется в виду то, что называется операционализацией понятий, — процедура исследования, устанавливающая соответствие между тем, *что* изучается, и тем, *как* изучается (она является относительно независимой от особенностей применения математических методов)[1]. Между тем, необходимо помнить, что качество любого исследования определяется прежде всего соответствием исходных данных той реальности, которая является предметом изучения. Если исследователь *понимает*, какое отношение имеют его данные к действительности (что они отражают), если он *уверен* в соответствии данных тому, что изучается и способен это *обосновать*, то ... Ответы на остальные вопросы исследования, я надеюсь, вы найдете в этой книге.

Я буду искренне признателен всем, кто сможет прислать свои предложения, пожелания, и главное — критические замечания по поводу этой книги.

Успехов!

А. Д. Н.
adn@an2806.spb.edu

[1] Некоторые издания по этой теме перечислены в разделе «Дополнительная литература» (Организация и планирование исследования).

ПСИХОЛОГИЯ И МАТЕМАТИКА

Более 200 лет назад великий И. Кант со свойственной ему убедительностью обосновывал несостоятельность психологии как науки исходя из того, что психические явления не поддаются измерению, а следовательно, к ним не применимы математические методы. Его соотечественник И. Гербарт противопоставил позиции И. Канта свою точку зрения в книге с названием «Психология как наука, заново обоснованная на опыте, метафизике и математике» (1824–1825). В ней он выражает свое мнение о связи психологии и математики: «Всякая теория, которая желает быть согласованной с опытом, прежде всего должна быть продолжена до тех пор, пока не примет количественных определений, которые являются в опыте или лежат в его основании. Не достигнув этого пункта, она висит в воздухе, подвергаясь всякому ветру сомнений и будучи неспособной вступить в связь с другими уже окрепшими воззрениями»[1]. Идеи И. Гербарта к концу XIX столетия воплощаются в жизнь отцами-основателями экспериментальной психологии. С тех пор *возможность* применения математических методов в психологии перестает вызывать сомнения. Но вопрос о *необходимости* их применения до сих пор вызывает дискуссии. Между тем проблема может быть решена признанием того, что психология — это и наука и искусство. Действительно, *искусству* практического консультирования или терапии вряд ли необходимо математическое обеспечение. Другое дело область *познания*, в том числе — того, что лежит в основе различных практических приемов. И здесь уже не достаточно обыденного понимания на уровне здравого смысла, необходим особый инструмент — научный метод, опирающийся на «количественные определения». Почему научное познание не довольствуется здравым смыслом, зачем необходимы математические методы?

Значение математических методов можно понять, сопоставляя обыденное и научное познание. На уровне обыденного познания действительности основным инструментом является здравый смысл. Результат познания — наше мнение (частное, субъективное). Мнение, или точка зрения по поводу той или иной проблемы, необходимо нам для прогноза или интерпретации грядущих реальных событий. Если прогнозы или интерпретации состоятельны, мы укрепляемся в своем мнении, если нет — мы вновь обращаемся к здравому смыслу и корректируем свое мнение, и т. д. Таким образом, продукт обыденного познания — мнение — прежде всего характеризуется как частное, субъек-

[1] Цитируется по кн.: *Корнилов К. Н.* Учение о реакциях человека с психологической точки зрения («Реактология»). М., 1923. С. 3.

тивное. И все мы хорошо знаем, насколько тяжело бывает переубедить другого человека или отстоять свое мнение. Произведение искусства — это тоже продукт обыденного познания, мнение творца, облеченное в специфическую форму. Эстетические переживания способствуют восприятию и принятию нами авторского мнения. Таким образом, обыденное познание, его продукт — мнение, его инструмент — здравый смысл лежат в основе наших представлений о действительности. А само понятие «обыденное» приобретает смысл в противовес альтернативному — «научному» познанию.

Научное познание по своей конечной цели — совершенству прогнозов и интерпретаций реальных событий — принципиально не отличается от обыденного познания. Более того, научное познание не отменяет и не заменяет обыденного, но добавляет кое-что для совершенствования его результатов — знаний и прогнозов. Наука стремится выйти за пределы частного мнения, сделать знания общезначимыми. В стремлении к общезначимости ученый обосновывает свое мнение *эмпирически, при помощи принятых в науке процедур*, возводя свое мнение в ранг научной теории. При этом предполагается (и практика это доказывает), что научное познание гарантирует нам более совершенные предсказания и интерпретации действительности.

Научное познание добавляет к инструменту обыденного познания — здравому смыслу — ряд дополнительных процедур, обеспечивая не только убедительность, но и *объективность* получаемых знаний. Рассмотрим их подробнее. Первый шаг любого (научного) исследования — выражение сомнения в истинности мнения, формулировка мнения как гипотезы — утверждения, допускающего проверку на фактах. Например, я могу поставить под сомнение свою точку зрения о том, что женщины более искусны в общении, чем мужчины. Но чтобы сделать гипотезу доступной проверке при помощи эмпирики, необходимо представить ее в форме математической модели, согласованной со способом регистрации наблюдений. Таким образом, гипотеза содержит указание на математическую модель, форма которой уточняется в соответствии с тем, как будет измерено то, что нас интересует. Моя содержательная гипотеза о большей искусности женщин в общении может быть представлена в форме математической модели: $M_м < M_ж$ (мужчины в среднем менее искусны в общении, чем женщины) или $f_м < f_ж$ (среди мужчин искусные в общении встречаются реже, чем среди женщин). В первом случае предполагается, что я могу вычислить среднюю «искусность в общении» для женщин и для мужчин по результатам ее количественного измерения при помощи некоторой специальной шкалы. Во втором случае достаточно определить частоту встречаемости «искусных в общении» среди мужчин и женщин.

Итак, научное познание начинается с нуждающегося в эмпирической проверке утверждения — гипотезы. Проверка гипотезы предполагает *измерение* интересующего исследователя явления и *обобщение* результатов измерения в виде, позволяющем сделать вывод в отношении гипотезы. Измерение и описание предполагает применение различных, хоть и взаимосвязанных, математических моделей и соответствующих им процедур. В процессе измерения мы представляем реальные события, явления, свойства в виде чисел, в соответствии

с принятой математической *моделью измерения*. Например, приписываем испытуемому число, обозначающее его пол (1 — мужской, 2 — женский), или ранг, соответствующий успешности выполнения задания (1 — лучше всех, 2 — второе место, и т. д.). Затем множество подобных результатов измерения мы должны представить в виде, доступном интерпретации с точки зрения выдвинутой гипотезы. Для этого используются математические *модели описания* для обобщения результатов измерения: менее сложные (частоты, средние значения и др.) или более сложные (корреляционный или факторный анализ и др.).

Помимо описания и измерения, существует и третье направление использования математики в психологии — *статистическая проверка гипотез*. Последнее направление тесно связано с общенаучными канонами экспериментального метода, основанными на статистическом выводе. Отдавая дань истории, отметим, что одним из первых примеров испытания статистической гипотезы была работа Дж. Арбутнота «Довод в пользу божественного провидения, выведенный из постоянной регулярности, наблюдаемой в рождении обоих полов» (1710–1712 гг.)[1]. Основываясь на том факте, что в течение 82 лет подряд мальчиков каждый год рождалось больше, чем девочек, автор показал, что эти данные опровергают гипотезу о равновероятном рождении мужчин и женщин. Если вероятность рождения мальчика точно равна 0,5, то вероятность того, что на протяжении 82 лет подряд мальчиков будет рождаться больше, чем девочек, равна $(^1/_2)^{82}$, т. е. она очень мала. По мнению Арбутнота, данный факт — результат вмешательства божественного Провидения, поскольку жизнь мужчины находится в большей опасности, чем жизнь женщины.

Общая логика статистической проверки гипотез, или определения статистической достоверности эмпирического результата, сохранилась в общих чертах и до настоящего времени. Возвращаясь к проверке моего мнения о женской искусности в общении, предположим, что я измерил ее при помощи 10-балльной шкалы у 32 женщин и 28 мужчин. Среднее значение для мужчин оказалось равным $M_м = 4{,}6$, а для женщин $M_ж = 5{,}1$. Здравый смысл мне подсказывает, что факт подтверждает мое мнение. Однако тут же возникает сомнение: достаточно ли столь малого различия в средних значениях, чтобы утверждать, что вообще *все* женщины в среднем более искусны в общении, чем *все* мужчины? Какова вероятность, что это все-таки не так? Для ответа на этот вопрос мне и необходимо обратиться к моделям статистического вывода. Если различия статистически значимы, то мое мнение приобретает статус научно обоснованного утверждения.

Таким образом, научное познание, в дополнение к здравому смыслу (но не вместо него!), обязательно предполагает применение математических методов, которые мы представили в виде трех классов моделей: измерения, описания и статистического вывода. Соотношение этих моделей в структуре познания схематично представлено на рис. 1.

Научное познание начинается с формулировки гипотезы — следствия теории или частного мнения по поводу некоторого аспекта реальности. Гипотеза

[1] *Кендалл М., Стьюарт А.* Статистические выводы и связи. М., 1973. С. 681.

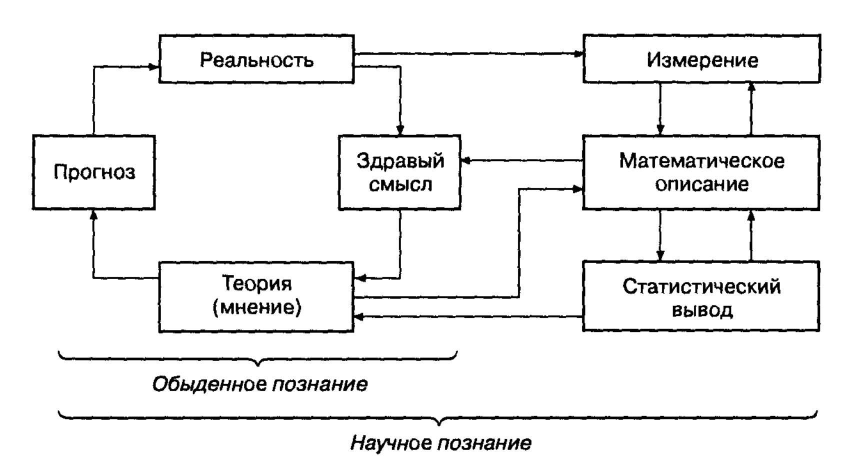


Рис. 1. Соотношение обыденного и научного познания

формулируется так, чтобы ее можно было проверить по результатам измерения, то есть в форме описательной математической модели. Описательная математическая модель согласуется с доступной измерительной моделью. Далее модель измерения применяется к интересующим нас аспектам действительности для регистрации результатов наблюдения (как правило — в числовой форме). Результаты измерения обобщаются при помощи описательной математической модели — для представления результатов измерения в доступном для интерпретации виде. Мы обращаемся к здравому смыслу и интерпретируем результаты применения описательных математических моделей. Однако чаще мы этим не ограничиваемся и обосновываем достоверность результатов при помощи соответствующей модели статистического вывода.

Изложенная логика аргументации характерна для науки в целом, в любых ее отраслях, в том числе для психологии. И гуманитарная специфика психологии вовсе не означает принципиального отличия научного метода психологии от методов других наук. Однако такая специфика предмета накладывает свой отпечаток на особенности применения математических методов. Это проявляется, в частности, в применяемых моделях измерения, в том, каким образом мы фиксируем результаты наблюдения непосредственно не видимого и не измеримого (способностей, тревожности и т. д.). Специфика измерительных моделей сказывается на применяемых описательных моделях, а те, в свою очередь — и на моделях статистического вывода.

Иногда можно слышать утверждения, что научный подход с применением математических методов необходим для академических научных исследований, а в практической работе вполне достаточно здравого смысла. Да, практическая деятельность психолога — это прежде всего *искусство* применения практических методов. Но здравого смысла недостаточно для профессиональной работы. Профессионал отличается тем, что может обосновать свою точку зрения, скажем, проверить эффективность того или иного практического метода или состоятельность организационного решения. При этом он будет опираться на научно обоснованные аргументы, а не только на собственное субъективное мнение.

Часть I

ОСНОВЫ ИЗМЕРЕНИЯ И КОЛИЧЕСТВЕННОГО ОПИСАНИЯ ДАННЫХ

Глава 1

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА

Исследование обычно начинается с некоторого предположения, требующего проверки с привлечением фактов. Это предположение — гипотеза — формулируется в отношении связи явлений или свойств в некоторой совокупности объектов.

ПРИМЕР

Исследователь может предположить, что женщины в среднем более тревожны, чем мужчины (тревожность связана с полом). Или что просмотр телепередач, содержащих сцены насилия, повышает агрессивность подростков. В первом случае исследователя интересуют такие явления, как тревожность и пол, а во втором — агрессивность и просмотр телепередач. Объектами-носителями свойств в первом случае будут все мужчины и женщины, а во втором — все подростки.

Для проверки подобных предположений на фактах необходимо измерить соответствующие свойства у их носителей. Но невозможно измерить тревожность у всех женщин и мужчин, как невозможно измерить агрессивность у всех подростков. Поэтому при проведении исследования ограничиваются лишь относительно небольшой группой представителей соответствующих совокупностей людей.

Генеральная совокупность — это все множество объектов, в отношении которого формулируется исследовательская гипотеза.

В первом примере такими генеральными совокупностями являются все мужчины и все женщины. Во втором — все подростки, которые смотрят телепередачи, содержащие сцены насилия. Генеральные совокупности, в отношении которых исследователь собирается сделать выводы по результатам исследования, могут быть по численности и более скромными.

ПРИМЕР

При изучении профессионального самоопределения студентов-выпускников некоторого факультета в конкретном вузе генеральная совокупность, казалось бы, весьма невелика и допускает сплошное исследование. Но исследователь обычно

надеется, что выводы исследования будут справедливы не только в отношении выпускников этого, но и последующих годов.

Таким образом, генеральная совокупность — это хотя и не бесконечное по численности, но, как правило, недоступное для сплошного исследования множество потенциальных испытуемых.

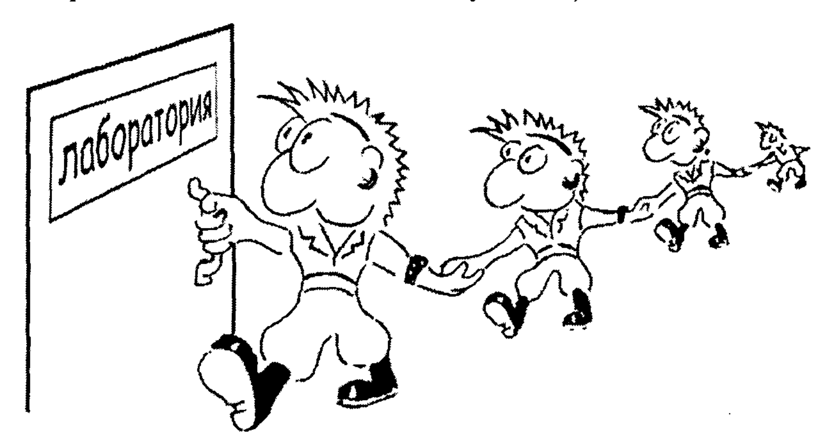
Выборка — это ограниченная по численности группа объектов (в психологии — испытуемых, респондентов), специально отбираемая из генеральной совокупности для изучения ее свойств. Соответственно, изучение на выборке свойств генеральной совокупности называется *выборочным исследованием.* Практически все психологические исследования являются выборочными, а их выводы распространяются на генеральные совокупности.

Таким образом, после того, как сформулирована гипотеза и определены соответствующие генеральные совокупности, перед исследователем возникает проблема организации выборки. Выборка должна быть такой, чтобы была обоснована генерализация выводов выборочного исследования — обобщение, распространение их на генеральную совокупность. *Основные критерии обоснованности выводов исследования — это репрезентативность выборки и статистическая достоверность (эмпирических) результатов.*

Репрезентативность выборки — иными словами, ее представительность — это способность выборки представлять изучаемые явления достаточно полно — с точки зрения их изменчивости в генеральной совокупности.

Конечно, полное представление об изучаемом явлении, во всем его диапазоне и нюансах изменчивости, может дать только генеральная совокупность. Поэтому репрезентативность всегда ограничена в той мере, в какой ограничена выборка. И именно репрезентативность выборки является основным критерием при определении границ генерализации выводов исследования. Тем не менее, существуют приемы, позволяющие получить достаточную для исследователя репрезентативность выборки.

Первый и основной прием — это **простой случайный (рандомизированный) отбор**. Он предполагает обеспечение таких условий, чтобы каждый член гене-



Случайный отбор?

ральной совокупности имел равные с другими шансы попасть в выборку. Случайный отбор обеспечивает возможность попадания в выборку самых разных представителей генеральной совокупности. При этом принимаются специальные меры, исключающие появление какой-либо закономерности при отборе. И это позволяет надеяться на то, что в конечном итоге в выборке изучаемое свойство будет представлено если и не во всем, то в максимально возможном его многообразии.

ПРИМЕР

Изучая агрессивность подростков, исследователь может *случайным* образом остановить свой выбор на 3 классах разных школ и затем *случайным* образом отобрать по 10 учащихся из каждого класса. Если же исследователь просит испытуемого пригласить на обследование своих друзей, он грубо нарушает принцип случайности отбора.

Второй способ обеспечения репрезентативности — это **стратифицированный случайный отбор**, или отбор по свойствам генеральной совокупности. Он предполагает предварительное определение тех качеств, которые могут влиять на изменчивость изучаемого свойства (это может быть пол, уровень дохода или образования и т. д.). Затем определяется процентное соотношение численности различающихся по этих качествам групп (страт) в генеральной совокупности и обеспечивается идентичное процентное соотношение соответствующих групп в выборке. Далее в каждую подгруппу выборки испытуемые подбираются по принципу простого случайного отбора.

ПРИМЕР

Исследователь резонно может предположить, что мальчики и девочки различаются как по агрессивности, так и по восприимчивости демонстрируемых по телевидению сцен насилия. Если исследователь планирует обобщить результат исследования влияния телевидения на агрессивность всех подростков, то, руководствуясь социально-демографическими данными, он должен обеспечить идентичное генеральной совокупности соотношение мальчиков и девочек в выборке.

Статистическая достоверность, или статистическая значимость, результатов исследования определяется при помощи методов статистического вывода. Эти методы мы будем подробно рассматривать во второй части этой книги. Сейчас лишь отметим, что они предъявляют определенные требования к численности, или *объему выборки.*

К сожалению, строгих рекомендаций по предварительному определению требуемого объема выборки не существует. Более того, ответ на вопрос о необходимой и достаточной ее численности исследователь обычно получает слишком поздно — только после анализа данных уже обследованной выборки. Тем не менее, можно сформулировать наиболее общие рекомендации:

- Наибольший объем выборки необходим при разработке диагностической методики — от 200 до 1000–2500 человек.

- Если необходимо сравнивать 2 выборки, их общая численность должна быть не менее 50 человек; численность сравниваемых выборок должна быть приблизительно одинаковой.
- Если изучается взаимосвязь между какими-либо свойствами, то объем выборки должен быть не меньше 30–35 человек.
- Чем больше *изменчивость* изучаемого свойства, тем больше должен быть объем выборки. Поэтому изменчивость можно уменьшить, увеличивая однородность выборки, например, по полу, возрасту и т. д. При этом, естественно, уменьшаются возможности генерализации выводов.

Зависимые и независимые выборки. Обычна ситуация исследования, когда интересующее исследователя свойство изучается на двух или более выборках с целью их дальнейшего сравнения. Эти выборки могут находиться в различных соотношениях — в зависимости от процедуры их организации. ***Независимые выборки*** характеризуются тем, что вероятность отбора любого испытуемого одной выборки не зависит от отбора любого из испытуемых другой выборки. Напротив, **зависимые выборки** характеризуются тем, что каждому испытуемому одной выборки поставлен в соответствие по определенному критерию испытуемый из другой выборки.

ПРИМЕР

Наиболее типичный пример зависимых выборок — повторное измерение свойства (свойств) на одной и той же выборке после воздействия (ситуация «до-после»). В этом случае выборки (одна — до, другая — после воздействия) зависимы в максимально возможной степени, так как они включают одних и тех же испытуемых. Могут быть и более слабые варианты зависимости. Например, мужья — одна выборка, их жены — другая выборка (при исследовании, например, их предпочтений). Или дети 5–7 лет — одна выборка, а их братья или сестры-близнецы — другая выборка.

В общем случае зависимые выборки предполагают попарный подбор испытуемых в сравниваемые выборки, а независимые выборки — независимый отбор испытуемых.

Следует отметить, что случаи «частично зависимых» (или «частично независимых») выборок недопустимы: это непредсказуемым образом нарушает их репрезентативность.

В заключение отметим, что можно выделить две парадигмы психологического исследования. Так называемая **R-методология** предполагает изучение изменчивости некоторого свойства (психологического) под влиянием некоторого воздействия, фактора либо другого свойства. *Выборкой является множество испытуемых.* Другой подход, **Q-методология**, предполагает исследование изменчивости субъекта (единичного) под влиянием различных стимулов (условий, ситуаций и т. д.). Ей соответствует ситуация, когда *выборкой является множество стимулов.*

Глава 2

ИЗМЕРЕНИЯ И ШКАЛЫ

ЧТО ТАКОЕ ИЗМЕРЕНИЕ

Любое эмпирическое научное исследование начинается с того, что исследователь фиксирует выраженность интересующего его свойства (или свойств) у объекта или объектов исследования, как правило при помощи чисел. Таким образом, следует различать *объекты исследования* (в психологии это чаще всего люди, испытуемые), их *свойства* (то, что интересует исследователя, составляет предмет изучения) и *признаки*, отражающие в числовой шкале выраженность свойств.

Измерение в терминах производимых исследователем операций — это приписывание объекту числа по определенному правилу. Это правило устанавливает соответствие между измеряемым свойством объекта и результатом измерения — признаком.

В обыденном сознании, как правило, нет необходимости разделять свойства вещей и их признаки: такие свойства предметов, как вес и длина, мы отождествляем, соответственно, с количеством граммов и сантиметров. Если нет необходимости в измерении, мы ограничиваемся сравнительными суждениями: этот человек тревожный, а этот — нет, этот более сообразителен, чем другой, и т. д.

В научном исследовании нам исключительно важно отдавать себе отчет в том, что точность, с которой признак отражает измеряемое свойство, зависит от процедуры (операции) измерения.

ПРИМЕР

Мы можем разделить всех наших испытуемых на две группы по сообразительности: сообразительные и не очень. И далее приписать каждому испытуемому символ (например, 1 и 0) в зависимости от его принадлежности к той или другой группе. А можем упорядочить всех испытуемых по степени выраженности сообразительности, приписывая каждому его ранг, от самого сообразительного (1 ранг), самого сообразительного из оставшихся (2 ранг) и т. д. до последнего испытуемого. В каком из этих двух случаев измеренный признак будет точнее отражать различия между испытуемыми по измеряемому свойству, догадаться нетрудно.

В зависимости от того, какая *операция* лежит в основе измерения признака, выделяют так называемые *измерительные шкалы*. Они еще называются шкалами С. Стивенса, по имени ученого-психолога, который их предложил. Эти шкалы устанавливают определенные соотношения между свойствами чисел и измеряемым свойством объектов. Шкалы разделяют на метрические (если есть или может быть установлена единица измерения) и неметрические (если единицы измерения не могут быть установлены).

ИЗМЕРИТЕЛЬНЫЕ ШКАЛЫ

Номинативная шкала (неметрическая), или шкала наименований (номинальное измерение). В ее основе лежит процедура, обычно не ассоциируемая с измерением. Пользуясь определенным правилом, объекты *группируются по различным классам так, чтобы внутри класса они были идентичны по измеряемому свойству*. Каждому классу дается наименование и обозначение, обычно числовое. Затем каждому объекту присваивается соответствующее обозначение.

ПРИМЕРЫ

Примеры номинативных признаков: «пол» (1 — мужской, 0 — женский), «национальность» (1 — русский, 2 — белорус, 3 — украинец), «предпочтение домашних животных» (1 — собаки, 2 — кошки, 3 — крысы, 0 — никакие) и т. д. В последнем случае если одному испытуемому присвоена 1, а другому 2, то это обозначает только то, что у них разные предпочтения: у первого — собаки, у второго — кошки. Из того, что 1 < 2, нельзя делать вывод, что у второго предпочтение выражено больше, чем у первого, и т. д.

Заметим, что в этом случае мы учитываем только одно свойство чисел — то, что это разные символы. Остальные свойства чисел не учитываются. Привычные операции с числами — упорядочивание, сложение-вычитание, деление — при измерении в номинативной шкале теряют смысл. При сравнении объектов мы можем делать вывод только о том, принадлежат они к одному или разным классам, тождественны или нет по измеренному свойству. Несмотря на такие ограничения, номинативные шкалы широко используются в психологии, и к ним применимы специальные процедуры обработки и анализа данных.

Ранговая, или **порядковая шкала (неметрическая)** (как результат ранжирования). Как следует из названия, измерение в этой шкале предполагает приписывание объектам чисел в зависимости от степени выраженности измеряемого свойства.

ПРИМЕР

Мы можем ранжировать всех испытуемых по интересующему нас свойству на основе экспертной оценки или по результатам выполнения некоторого задания и при-

писать каждому испытуемому его ранг. Или предложить испытуемым самим определить выраженность изучаемого свойства, пользуясь предложенной шкалой (5-, 7- или 10-балльной).

Существует множество способов получения измерения в порядковой шкале. Но суть остается общей: при сравнении испытуемых друг с другом мы можем сказать, больше или меньше выражено свойство, но не можем сказать, насколько больше или насколько меньше оно выражено, а уж тем более — во сколько раз бо́льше или меньше. При измерении в ранговой шкале, таким образом, из всех свойств чисел учитывается то, что они разные, и то, что одно число больше, чем другое.

ПРИМЕР

Четверым бегунам присвоены ранги в соответствии с тем, кто раньше достиг «финиша» (ранг 1 — самый быстрый):

Бегун	Ранг
A	1
B	2
C	3
D	4

Основываясь только на этих данных, мы можем судить о том, кто раньше прибежал, а кто позже. Но мы не можем судить, *насколько* каждый из них пробежал быстрее или медленнее другого. Глядя на эти ранги, можно было бы предположить, что бегуны A и B различаются меньше, чем бегуны B и D, так как 2–1 = 1, а 4–2 = 2. Однако такой вывод — следствие «пленяющей магии чисел»: бегун A мог быть тренированным спортсменом, пробежавшим дистанцию в 2 раза быстрее, чем бегуны B, C и D — «увальни», пришедшие к «финишу» с минимальными различиями во времени.

При ранжировании «вручную», а не при помощи компьютера, следует иметь в виду два обстоятельства:

1. Установите для себя и запомните порядок ранжирования. Вы можете ранжировать испытуемых по их «месту в группе»: ранг 1 присваивается тому, у которого наименьшая выраженность признака, и далее — увеличение ранга по мере увеличения уровня признака. Или можно ранг 1 присваивать тому, у которого 1-е место по выраженности данного признака (например, «самый быстрый»). Строгих правил выбора здесь нет, но важно помнить, в каком направлении производилось ранжирование.

2. Соблюдайте правило ранжирования для связанных рангов, когда двое или более испытуемых имеют одинаковую выраженность измеряемого свойства. В этом случае таким испытуемым присваивается один и тот же, средний ранг. Например, если вы ранжируете испытуемых по «месту в группе» и двое имеют одинаковые самые высокие исходные оценки, то обоим присваивается средний ранг 1,5: $(1+2)/2 = 1,5$. Следующему за этой парой испытуемому присваивается ранг 3, и т. д. Это правило основано на соглашении соблюде-

ния одинаковой суммы рангов для связанных и несвязанных рангов. *В соответствии с этим правилом сумма всех присвоенных рангов для группы численностью N должна равняться N(N+1)/2, вне зависимости от наличия или отсутствия связей в рангах.*

Интервальная шкала (метрическая). Это такое измерение, при котором числа отражают не только различия между объектами в уровне выраженности свойства (характеристика порядковой шкалы), но и то, насколько больше или меньше выражено свойство. Равным разностям между числами в этой шкале соответствуют равные разности в уровне выраженности измеренного свойства. Иначе говоря, измерение в этой шкале предполагает возможность применения *единицы измерения* (*метрики*). Объекту присваивается число единиц измерения, пропорциональное выраженности измеряемого свойства. Важная особенность интервальной шкалы — произвольность выбора нулевой точки: ноль вовсе не соответствует полному отсутствию измеряемого свойства. Произвольность выбора нулевой точки отсчета обозначает, что измерение в этой шкале не соответствует *абсолютному* количеству измеряемого свойства. Следовательно, применяя эту шкалу, мы можем судить, насколько больше или насколько меньше выражено свойство при сравнении объектов, но не можем судить о том, во сколько раз больше или меньше выражено свойство.

ПРИМЕР

Наиболее типичный пример измерения в интервальной шкале — температура по шкале Цельсия (°С). Важная особенность такого измерения заключается в том, что нулевая точка на шкале не соответствует полному отсутствию измеряемого свойства (0 °С — это точка замерзания воды, но не отсутствия температуры, тепла). И если сегодня +5 °С, а вчера было +10 °С, то можно сказать, что сегодня на 5 градусов холоднее, но неверно утверждать, что сегодня холоднее в два раза.



Интервальные измерения широко используются в психологии. Примером могут являться тестовые шкалы, которые специально вводятся при обосновании равноинтервальности (метричности) тестовой шкалы (IQ Векслера, стены, *Т*-шкала и т. д.).

Абсолютная шкала, или шкала отношений (метрическая). Измерение в этой шкале отличается от интервального только тем, что в ней устанавливается нулевая точка, соответствующая полному отсутствию выраженности измеряемого свойства.

ПРИМЕР

В отличие от температуры по Цельсию, температура по Кельвину представляет собой измерение в абсолютной шкале. Более привычные примеры измерения в этой шкале — это измерения роста, веса, времени выполнения задачи и т. д. Общим в этих примерах является применение единиц измерения и то, что нулевой точке соответствует полное отсутствие измеряемого свойства.

В силу абсолютности нулевой точки, при сравнении объектов мы можем сказать не только о том, насколько больше или меньше выражено свойство, но и о том, во сколько раз (на сколько процентов и т. д.) больше или меньше оно выражено. Измерив время решения задачи парой испытуемых, мы можем сказать не только о том, кто и на сколько секунд (минут) решил задачу быстрее, но и о том, во сколько раз (на сколько процентов) быстрее.

Следует отметить, что, несмотря на привычность и обыденность абсолютной шкалы, в психологии она используется не часто. Из редких примеров можно привести измерение времени реакции (обычно в миллисекундах) и измерение абсолютных порогов чувствительности (в физических единицах свойств стимула).

Перечисленные шкалы полезно характеризовать еще и по признаку их дифференцирующей способности (*мощности*). В этом отношении шкалы по мере возрастания мощности располагаются следующим образом: номинативная, ранговая, интервальная, абсолютная. Таким образом, неметрические шкалы заведомо менее мощные — они отражают меньше информации о различии объектов (испытуемых) по измеренному свойству, и, напротив, метрические шкалы более мощные, они лучше дифференцируют испытуемых. Поэтому, если у исследователя есть возможность выбора, следует применить более мощную шкалу. Другое дело, что чаще такого выбора нет, и приходится использовать доступную измерительную шкалу. Более того, часто исследователю даже трудно определить, какую шкалу он применяет.

КАК ОПРЕДЕЛИТЬ, В КАКОЙ ШКАЛЕ ИЗМЕРЕНО ЯВЛЕНИЕ

Определение того, в какой шкале измерено явление (представлен признак), — ключевой момент анализа данных: любой последующий шаг, выбор любого метода зависит именно от этого.

Обычно идентификация номинативной шкалы, ее дифференциация от ранговой, а тем более от метрической шкалы не вызывает особых проблем.

ПРИМЕР

Рассмотрим вопрос анкеты, для ответа на который испытуемые выбирают один из предложенных вариантов:

«Насколько Вы уверены в своих силах...

1) Совершенно уверен
2) Затрудняюсь ответить
3) Совершенно неуверен»

Если исследователя интересует, в какой степени испытуемые уверены или не уверены в своих силах, то логично предполагать, что признак представлен в ранговой шкале. Если же исследователя интересует то, как распределились ответы по вариантам или чем характеризуется каждая из 3 соответствующих групп, то разумнее рассматривать этот признак как номинативный.

Значительно сложнее определить различие между порядковой и метрической шкалами. Проблема связана с тем, что измерения в психологии, как правило, косвенные. Непосредственно мы измеряем некоторые наблюдаемые явления или события: количество ответов на вопросы, или заданий, решенных за отведенное время, или время решения набора заданий и т. д. Но при этом выносим суждения о некотором скрытом, латентном свойстве, недоступном прямому наблюдению: об агрессивности, общительности, способности и т. д.

Количество заданий, решенных за отведенное время, — это, конечно, измерение в метрической шкале. Но само по себе это количество нас интересует лишь в той мере, в какой оно отражает некоторую изучаемую нами способность. Соответствуют ли равные разности решенных задач равным разностям выраженности изучаемого свойства (способности)? Если ответ «да» — шкала метрическая (интервальная), если «нет» — шкала порядковая.

Конечно, проще всего в подобных ситуациях согласиться с тем, что признак представлен в порядковой шкале. Но при этом мы существенно ограничиваем себя в выборе методов последующего анализа. Более того, переход к менее мощной шкале обрекает нас на утрату части столь ценной для нас эмпирической информации об индивидуальных различиях испытуемых. Следствием этого может являться падение статистической достоверности результатов исследования. Поэтому исследователь стремится все же найти свидетельства того, что используемая шкала — более мощная, метрическая. То, какие обоснования метричности шкалы обычно учитываются, мы рассмотрим несколько позднее — в разделе о нормальном распределении.

Задачи и упражнения

Определите, в какой шкале представлено каждое из приведенных ниже измерений: наименований, порядка, интервалов, абсолютной.

1. Порядковый номер испытуемого в списке (для его идентификации).
2. Количество вопросов в анкете как мера трудоемкости опроса.
3. Упорядочивание испытуемых по времени решения тестовой задачи.
4. Академический статус (ассистент, доцент, профессор) как указание на принадлежность к соответствующей категории.
5. Академический статус (ассистент, доцент, профессор) как мера продвижения по службе.
6. Телефонные номера.
7. Время решения задачи.
8. Количество агрессивных реакций за рабочий день.
9. Количество агрессивных реакций за рабочий день как показатель агрессивности.

Глава 3

ТАБЛИЦЫ И ГРАФИКИ

ТАБЛИЦА ИСХОДНЫХ ДАННЫХ

Обычно в ходе исследования интересующий исследователя признак измеряется не у одного-двух, а у множества объектов (испытуемых). Кроме того, каждый объект характеризуется не одним, а целым рядом признаков, измеренных в разных шкалах. Одни признаки представлены в номинативной шкале и указывают на принадлежность испытуемых к той или иной группе (пол, профессия, контрольная или экспериментальная группа и т. д.). Другие признаки могут быть представлены в порядковой или метрической шкале. Поэтому результаты измерения для дальнейшего анализа чаще всего представляют в виде *таблицы исходных данных*. Каждая строка такой таблицы обычно соответствует одному *объекту*, а каждый столбец — одному измеренному *признаку*. Таким образом, исходной формой представления данных является таблица типа «объект — признак». В ходе дальнейшего анализа каждый признак выступает в качестве переменной величины, или просто — *переменной*, значения которой меняются от объекта к объекту.

ПРИМЕР

Предположим, психолога интересует социальная сплоченность двух параллельных классов, различие в этом отношении мальчиков и девочек и эффективность проведенного в одном из этих классов социально-психологического тренинга. Для измерения социальной сплоченности исследователь задавал каждому ученику до и после тренинга один и тот же вопрос: «Как часто твое мнение совпадает с мнением твоих одноклассников?». Для ответа ученикам предлагалось выбрать один из пяти вариантов: 1 — никогда, 2 — редко, 3 — затрудняюсь ответить, 4 — часто, 5 — всегда. Исходные данные исследования представлены в табл. 3.1.

Общая численность всех испытуемых $N = 60$. Численность класса, с которым проводился тренинг, $N_1 = 30$; численность другого класса — $N_2 = 30$. Первые два столбца таблицы — порядковый номер испытуемого (№) и Ф. И. О. Далее следуют четыре столбца, соответствующие четырем интересующим исследователя признакам:

x_{1i} — пол (номинативный),

x_{2i} — класс (номинативный),

Таблица 3.1

Таблица исходных данных

1	2	3	4	5	
№	Ф.И.О.	Пол	Класс	Самооценка	
				До	После
		1	2	3	4
1	Иванов И. О.	1	0	5	5
2	Васильев К. А.	1	1	3	4
3	Розова М. И.	0	1	2	3
4	Краснова О. С.	0	0	3	3
5	Цветов С. Т.	1	0	1	3
6	Лозовая Е. И.	0	1	4	4
...	...	...	...	...	...
i	...	x_{1i}	x_{2i}	x_{3i}	x_{4i}
...	...	...	...	...	...
60	Петров Е. М.	1	1	3	3

x_{3i} — самооценка до тренинга (порядковый),
x_{4i} — самооценка после тренинга (порядковый),
где i — текущий номер испытуемого (меняется от 1 до $N = 60$).

Обратите внимание на то, что нумерация испытуемых в таблице исходных данных (табл. 3.1) — сквозная, вне зависимости от принадлежности к той или иной группе. То, к какой группе принадлежит испытуемый, определяется значением соответствующей номинативной переменной (пол, класс).

ТАБЛИЦЫ И ГРАФИКИ РАСПРЕДЕЛЕНИЯ ЧАСТОТ

Как правило, анализ данных начинается с изучения того, как часто встречаются те или иные значения интересующего исследователя признака (переменной) в имеющемся множестве наблюдений. Для этого строятся *таблицы и графики распределения частот*. Нередко они являются основой для получения ценных содержательных выводов исследования.

Если признак принимает всего лишь несколько возможных значений (до 10–15), то таблица распределения частот показывает частоту встречаемости каждого значения признака. Если указывается, сколько раз встречается каждое значение признака, то это — таблица *абсолютных* частот распределения, если указывается доля наблюдений, приходящихся на то или иное значение признака, то говорят об *относительных* частотах распределения.

ПРИМЕР

Предположим, исследователя в нашем примере (табл. 3.1) интересует, как распределяются ответы всех учеников до проведения тренинга. Для этого он подсчитает частоту встречаемости каждого из ответов и составит таблицу распределения частот (табл. 3.2). Таблица показывает, что чаще встречаются средние значения выраженности признака и реже — крайние значения.

Таблица 3.2

Таблица распределения частот

Значение	f_a (абсолютная частота)	f_o (относительная частота)	f_{cum} (накопленная частота)
5	3	0,05	1,00
4	12	0,20	0,95
3	21	0,35	0,75
2	15	0,25	0,40
1	9	0,15	0,15
Σ (сумма):	60	1	—

Абсолютная и относительная частоты связаны соотношением:

$$f_o = \frac{f_a}{N}, \qquad (3.1)$$

где f_a — абсолютная частота некоторого значения признака, N — число наблюдений, f_o — относительная частота этого значения признака. Очевидно, что сумма всех абсолютных частот равна числу наблюдений — N, а сумма всех относительных частот равна 1. Нередко относительная частота применяется для оценки вероятности встречаемости значения.

Во многих случаях признак может принимать множество различных значений, например, если мы измеряем время решения тестовой задачи. В этом случае о распределении признака позволяет судить *таблица сгруппированных частот*, в которых частоты группируются по разрядам или интервалам значений признака.

ПРИМЕР

Предположим, в группе испытуемых численностью 40 человек измерено время решения тестовой задачи. Максимальное время составило 67 секунд, минимальное — 32 секунды. Построение таблицы распределения частот в этом случае производится поэтапно.

Построение таблицы сгруппированных частот.

1. Определение размаха: 67–32 = 35.
2. Выбор желаемого числа разрядов и интервала разрядов. Определяется произвольно. Обычное число разрядов — от 6 до 15. Удобным интервалом разрядов в нашем случае может быть 5. 35 делим на 5, получаем число разрядов — 7. Учитывая, что начинать лучше с 30 или с 31 и заканчивать на 69 или 70, уточняем размах (70–30 = 40) и число разрядов (40/5 = 8).

3. Определение границ разрядов. Если мы начнем с 30, то первый разряд будет с 30 до 34, второй — с 35 до 49 и т. д., до восьмого — с 65 до 69. Границы соседних разрядов не должны совпадать!
4. Подсчет частот встречаемости значений признака для каждого интервала.

Табл. 3.3 содержит результат подсчета сгруппированных таким образом частот по разрядам (интервалам) значений признака — времени решения тестовой задачи.

Таблица 3.3

Таблица частот, сгруппированных по интервалам времени решения тестовой задачи

Интервал времени, с	f_{aj} (абсолютная частота)	f_{oj} (относительная частота)	f_{cum} (накопленная частота)
30–34	1	0,025	0,025
35–39	2	0,050	0,075
40–44	5	0,125	0,200
45–49	8	0,200	0,400
50–54	10	0,250	0,650
55–59	8	0,200	0,850
60–64	4	0,100	0,950
65–69	2	0,050	1,000
Σ (сумма):	40	1,000	—

Еще одной разновидностью таблиц распределения являются таблицы распределения *накопленных* частот. Они показывают, как накапливаются частоты по мере возрастания значений признака. Напротив каждого значения (интервала) указывается сумма частот встречаемости всех тех наблюдений, величина признака у которых не превышает данного значения (меньше верхней границы данного интервала). Накопленные частоты содержатся в правых столбцах табл. 3.2 и 3.3.

Для более наглядного представления строится график распределения частот или график накопленных частот — гистограмма или сглаженная кривая распределения.

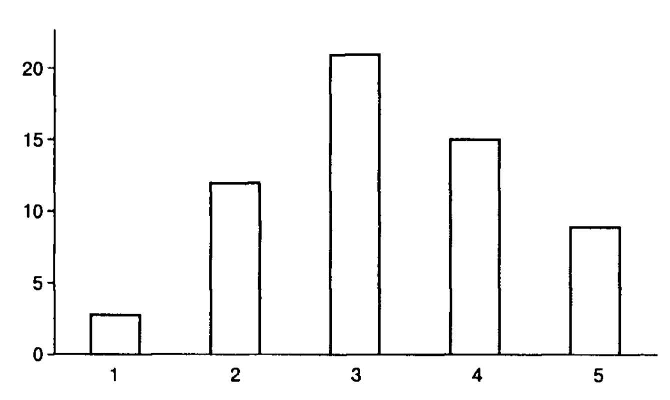


Рис. 3.1. Гистограмма распределения частот самооценки (по данным табл. 3.2)

Гистограмма распределения частот — это столбиковая диаграмма, каждый столбец которой опирается на конкретное значение признака или разрядный интервал (для сгруппированных частот). Высота столбика пропорциональна частоте встречаемости соответствующего значения. На рис. 3.1 изображена гистограмма распределения частот для примера из табл. 3.2.

Гистограмма накопленных частот отличается от гистограммы распределения тем, что высота каждого столбика пропорциональна частоте, накопленной к данному значению (интервалу). На рис. 3.2 изображена гистограмма накопленных частот для данных табл. 3.2.

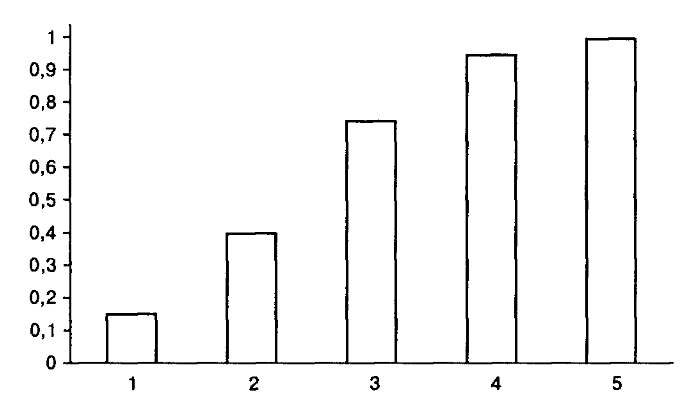


Рис. 3.2. Гистограмма накопленных относительных частот самооценки (по данным табл. 3.2)

Построение *полигона распределения частот* напоминает построение гистограммы. В гистограмме вершина каждого столбца, соответствующая частоте встречаемости данного значения (интервала) признака, — отрезок прямой. А для полигона отмечается точка, соответствующая середине этого отрезка. Далее все точки соединяются ломаной линией (рис. 3.3).

Вместо гистограммы или полигона часто изображают сглаженную кривую распределения частот. На рис. 3.4 изображена гистограмма распределения для примера из табл. 3.3 (столбики) и сглаженная кривая того же распределения частот.

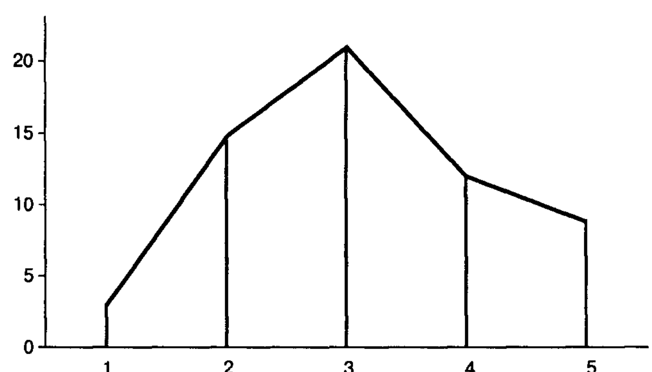


Рис. 3.3. Полигон распределения частот самооценки (по данным табл. 3.2)

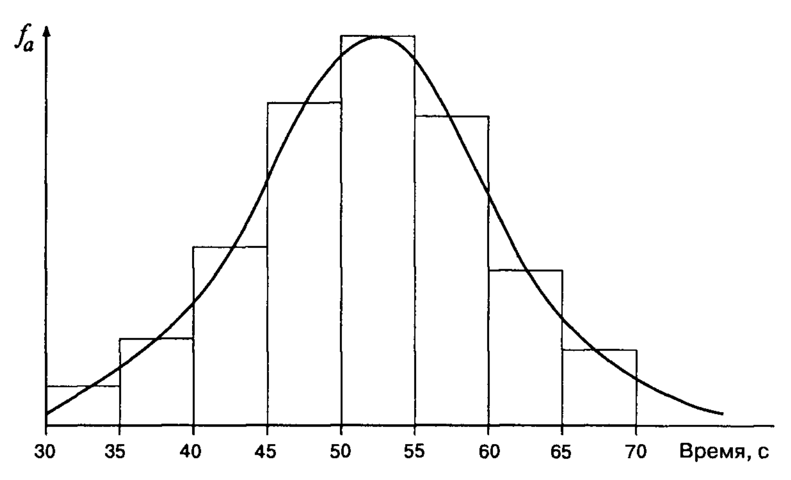


Рис. 3.4. Гистограмма и сглаженный график распределения частот времени решения тестовой задачи (по данным табл. 3.3)

ПРИМЕНЕНИЕ ТАБЛИЦ И ГРАФИКОВ РАСПРЕДЕЛЕНИЯ ЧАСТОТ

Таблицы и графики распределения частот дают важную предварительную информацию о *форме распределения признака*: о том, какие значения встречаются реже, а какие чаще, насколько выражена изменчивость признака. Обычно выделяют следующие типичные формы распределения. *Равномерное распределение* — когда все значения встречаются одинаково (или почти одинаково) часто. *Симметричное распределение* — когда одинаково часто встречаются крайние значения. *Нормальное распределение* — симметричное распределение, у которого крайние значения встречаются редко и частота постепенно повышается от крайних к серединным значениям признака. *Асимметричные распределения* — *левосторонние* (с преобладанием частот малых значений), *правосторонние* (с преобладанием частот больших значений). К понятию формы распределения мы еще не раз вернемся, прежде всего — в связи с использованием в психологии нормального распределения как особого эталона — стандарта.

Уже сами по себе таблицы и графики распределения признака позволяют делать некоторые содержательные выводы при сравнении групп испытуемых между собой. Сравнивая распределения, мы можем не только судить о том, какие значения встречаются чаще в той или иной группе, но и сравнивать группы по степени выраженности индивидуальных различий — *изменчивости* по данному признаку.

Таблицы и графики накопленных частот позволяют быстро получить дополнительную информацию о том, сколько испытуемых (или какая их доля) имеют выраженность признака не выше определенного значения.

Следует отметить, что для сравнения групп разной численности следует использовать таблицы и графики относительных частот.

ПРИМЕР

В группе юношей и группе девушек измерена тревожность при помощи тестовой шкалы. По результатам измерений построены сглаженные графики распределения относительных частот отдельно для юношей и девушек (рис. 3.5). Сравнивая графики, можно сделать содержательные выводы как по уровню выраженности, так и по индивидуальной изменчивости тревожности у юношей и девушек. Так, юноши в среднем менее тревожны, чем девушки. Но индивидуальные различия — изменчивость — по тревожности выше у юношей, чем у девушек: девушки в этом отношении более похожи друг на друга.

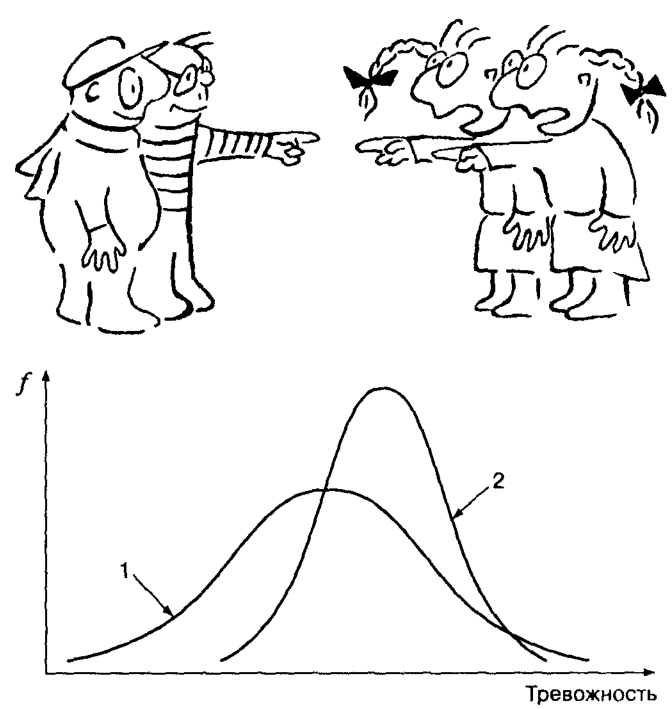


Рис. 3.5. Графики распределения относительных частот тревожности юношей (1) и девушек (2)

ТАБЛИЦЫ СОПРЯЖЕННОСТИ НОМИНАТИВНЫХ ПРИЗНАКОВ

Таблицы сопряженности, или кросстабуляции — это таблицы совместного распределения частот двух и более номинативных признаков, измеренных на одной группе объектов. Эти таблицы позволяют сопоставить два или более

распределения. Столбцы такой таблицы соответствуют категориям (градациям) одного номинативного признака, а строки — категориям (градациям) другого номинативного признака. Если номинативные признаки внесены в электронную таблицу исходных данных, то таблицу сопряженности можно построить, воспользовавшись функцией «Кросстабуляция» одного из стандартных статистических пакетов (например, Crosstabs — в SPSS).

ПРИМЕР

В одном из исследований изучалась склонность людей передавать плохие или хорошие новости. На ветровых стеклах автомобилей, припаркованных у почтовых ящиков, были оставлены почтовые открытки с указанием адресата (всего — 180 шт.), содержащие либо нейтральные (хорошие), либо плохие новости. В качестве плохой новости использовалось сообщение о супружеской неверности супруга (супруги) — получателя сообщения. В процессе исследования подсчитывалось количество отправленных открыток, дошедших до указанного адреса. Результаты представлены в табл. 3.4 — в виде таблицы сопряженности частот двух номинативных признаков: новость (две градации: плохая — хорошая), сообщение (две градации: отправлено — не отправлено). Как видите, таблица дает основание делать вывод о том, что люди с меньшей охотой отправляли открытки, содержащие плохие новости.

Таблица 3.4

Зависимость распределения оставленных и полученных открыток от их содержания

		Сообщения	
		отправлено	не отправлено
Новость	Хорошая	35	25
	Плохая	23	97
Σ		58	122

Конечно, таблицы сопряженности могут включать номинативные признаки, имеющие и более двух градаций. Например, по табл. 3.1 для изучения различий в самооценке мальчиков и девочек исследователь мог бы построить таблицу сопряженности признаков «Пол» (две градации) и «Самооценка» (пять градаций).

Задачи и упражнения

На трех разных, достаточно больших группах испытуемых изучалась диагностическая ценность методики измерения креативности. Методика представляла собой 10 заданий, которые испытуемые решали за определенный промежуток времени. Фиксировалось количество решенных заданий (минимум — 0, максимум — 10). По результатам исследования была построена табл. 3.5, позволяющая сравнить три группы по распределению относительных частот (в процентах) показателей креативности.

Таблица 3.5

Таблица распределения результатов измерения креативности в трех группах

Решенные задания	Относительные частоты (%)		
	группа 1	группа 2	группа 3
0	1	10	0
1	4	20	0
2	5	30	1
3	10	30	2
4	20	5	3
5	30	3	4
6	20	1	10
7	5	0	15
8	3	0	25
9	1	0	25
10	1	0	15

1. Для какой из групп задания были слишком легкие, а для какой — слишком трудные?
2. В какой группе наблюдается наибольшая, а в какой — наименьшая индивидуальная изменчивость результатов?
3. В отношении какой группы, на ваш взгляд, методика может иметь наибольшую диагностическую ценность — точнее измерять индивидуальные различия?

ОБРАБОТКА НА КОМПЬЮТЕРЕ

1. Таблица исходных данных. Может быть образована в среде SPSS двумя способами. А) Данные можно предварительно набрать в среде программы Excel (строки — испытуемые, столбцы — признаки). Затем путем простого копирования блока данных в таблице Excel перенести при помощи команды «вставка» (**Past...**) этот блок данных в предварительно открытую пустую таблицу SPSS и сохранить ее. Б) Данные можно набирать сразу в программе SPSS. Полезно затем каждой переменной присвоить имя, вместо принятого в SPSS по умолчанию (`var0001`...). Начиная пользоваться программой SPSS, убедитесь, что в качестве разделителя целой и дробной частей установлен единый символ для всех программ — точка (**Панель управления > Языки и стандарты > Числа > Разделитель целой и дробной частей числа** — установить точку)!

2. Таблицы распределения частот. Выбираем **Analyze > Descriptive Statistics > Frequencies...** В открывшемся диалоговом окне (**Frequencies**) переносим из левой в правую часть интересующие нас переменные. После этого нажимаем ОК. В окне результатов (**Output...**) для каждой переменной получаем таблицу

распределения с предварительным указанием объема выборки (**Valid**) и числа пропущенных значений (**Missing**). В таблице распределения каждая строка соответствует отдельному значению, для которого указаны (столбцы): абсолютная частота (**Frequency**), относительная частота в процентах от объема выборки — без учета пропусков (**Percent**), относительная частота действительного числа наблюдений — с учетом пропусков (**Valid Percent**), накопленная относительная частота в процентах (**Cumulative Percent**).

3. Графики распределения частот. А) При построении таблиц распределения частот (см. предыдущий пункт) в открывшемся диалоговом окне после выбора переменных нажать кнопку **Charts...** (графики). Задать тип графика (**Chart Type**) — гистограммы (**Histograms**). Нажать **Continue**, затем ОК. Вместе с таблицей распределения частот вы получите гистограмму распределения каждого выбранного признака. Б) Выбираем **Graphs > Histogram...** В открывшемся диалоговом окне переносим из левой в правую часть интересующую нас переменную, нажимаем ОК. Получаем гистограмму распределения этой переменной.

4. Таблицы сопряженности (кросстабуляции). Выбираем **Analyze > Descriptive Statistics > Crosstabs...** В открывшемся окне диалога выбираем интересующие нас номинативные переменные: одну для строк (**Row(s)**), другую — для столбцов (**Column(s)**). После нажатия ОК получаем таблицу кросстабуляции (сопряженности) в абсолютных значениях частот. Если в окне диалога нажать кнопку **Cells...** (Ячейки), то в открывшемся окне можно установкой флажков задать вывод относительных частот в процентах (**Percentages**) по строкам (**Row**), столбцам (**Columns**) или в целом по таблице (**Total**).

Глава 4

ПЕРВИЧНЫЕ ОПИСАТЕЛЬНЫЕ СТАТИСТИКИ

К первичным описательным статистикам (*Descriptive Statistics*) обычно относят числовые характеристики распределения измеренного на выборке признака. Каждая такая характеристика отражает *в одном числовом значении* свойство распределения *множества результатов измерения*: с точки зрения их *расположения* на числовой оси либо с точки зрения их *изменчивости*. Основное назначение каждой из первичных описательных статистик — замена множества значений признака, измеренного на выборке, одним числом (например, средним значением как мерой центральной тенденции). Компактное описание группы при помощи первичных статистик позволяет интерпретировать результаты измерений, в частности, путем сравнения первичных статистик разных групп.

МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ

Мера центральной тенденции (*Central Tendency*) — это число, характеризующее выборку по уровню выраженности измеренного признака.

Существуют три способа определения «центральной тенденции», каждому из которых соответствует своя мера: мода, медиана и выборочное среднее.

Мода (*Mode*) — это такое значение из множества измерений, которое встречается наиболее часто. Моде, или *модальному интервалу* признака, соответствует наибольший подъем (вершина) графика распределения частот. Если график распределения частот имеет одну вершину, то такое распределение называется *унимодальным*.

ПРИМЕР

Среди 8 значений признака (3, 7, 3, 5, 7, 8, 7, 6) мода Mo = 7 как наиболее часто встречающееся значение. В табл. 3.2 предыдущего параграфа Mo = 3, а в табл. 3.3 модальным является интервал 50–54.

Когда два соседних значения встречаются одинаково часто и чаще, чем любое другое значение, мода есть среднее этих двух значений.

Распределение может иметь и не одну моду. Когда все значения встречаются одинаково часто, принято считать, что такое распределение не имеет моды.

Бимодальное распределение имеет на графике распределения две вершины, даже если частоты для двух вершин не строго равны. В последнем случае выделяют большую и меньшую моду. Во всей группе может быть и несколько локальных вершин распределения частот. Тогда выделяют *наибольшую моду* и *локальные моды.*

Еще раз отметим, что мода — это *значение* признака, а не его частота.

Медиана (*Median*) — это такое значение признака, которое делит упорядоченное (ранжированное) множество данных пополам так, что одна половина всех значений оказывается меньше медианы, а другая — больше. Таким образом, первым шагом при определении медианы является упорядочивание (ранжирование) всех значений по возрастанию или убыванию. Далее медиана определяется следующим образом:

- если данные содержат нечетное число значений (8, 9, 10, 13, 15), то медиана есть центральное значение, т. е. $Md = 10$;
- если данные содержат четное число значений (5, 8, 9, 11), то медиана есть точка, лежащая посередине между двумя центральными значениями, т. е. $Md = (8+9)/2 = 8{,}5$.

Среднее (*Mean*) (M_x — выборочное среднее, среднее арифметическое) — определяется как сумма всех значений измеренного признака, деленная на количество суммированных значений.

Если некоторый признак X измерен в группе испытуемых численностью N, мы получим значения: $x_1, x_2, ..., x_i, ..., x_N$ (где i — текущий номер испытуемого, от 1 до N). Тогда среднее значение M_x определяется по формуле:

$$M_x = \frac{1}{N}\sum_{i=1}^{N} x_i \,. \qquad (4.1)$$

Свойства среднего. Если к каждому значению переменной прибавить одно и то же число c, то среднее увеличится на это число (уменьшится на это число, если оно отрицательное):

$$M_{(x_i+c)} = \frac{1}{N}\sum_{i=1}^{N}(x_i + c) = M_x + c \,. \qquad (4.2)$$

А если каждое значение переменной умножить на одно и то же число c, то среднее увеличится в c раз (уменьшится в c раз, если делить на c):

$$M_{(x_i \cdot c)} = \frac{1}{N}\sum_{i=1}^{N}(x_i \cdot c) = M_x \cdot c \,. \qquad (4.3)$$

Далее мы неоднократно будем обращаться к такой величине, как *отклонение от среднего*: $(x_i - M_x)$. Из первого, очевидного свойства среднего следует

еще одно важное свойство, не столь очевидное: *сумма всех отклонений от среднего равна нулю:*

$$\sum_{i=1}^{N}(x_i - M_x) = 0\,. \tag{4.4}$$

Соответственно, среднее отклонение от среднего также равно 0.

ВЫБОР МЕРЫ ЦЕНТРАЛЬНОЙ ТЕНДЕНЦИИ

Каждая мера центральной тенденции обладает характеристиками, которые делают ее ценной в определенных условиях.

Для *номинативных* данных, разумеется, единственной подходящей мерой центральной тенденции является мода, или *модальная категория — та градация номинативной переменной, которая встречается наиболее часто.*

Для порядковых и метрических переменных, распределение которых унимодальное и симметричное, мода, медиана и среднее совпадают. Чем больше отклонение от симметричности, тем больше расхождение между значениями этих мер центральной тенденции. По этому расхождению можно судить о том, насколько симметрично или асимметрично распределение.

Наиболее очевидной и часто используемой мерой центральной тенденции является среднее значение. Но его использование ограничивается тем, что *на величину среднего влияет каждое отдельное значение.* Если какое-нибудь значение в группе увеличится на c, то среднее увеличится на c/N. Таким образом, среднее значение весьма чувствительно к «выбросам» — экстремально малым или большим значениям переменной.

На величину моды и медианы величина каждого отдельного значения не влияет. Например, если в группе из 20 измерений переменной наибольшее значение утроится по величине, то не изменится ни мода, ни медиана. Величина среднего при этом заметно изменится. Иначе говоря, мода и медиана не чувствительны к «выбросам».



ПРИМЕР

Если 9 человек имеют месячный доход от 5000 до 6000 рублей, со средним 5600 рублей, а доход десятого составляет 15000 рублей, то средний доход для этих 10 человек составит 6540 рублей. Эта цифра не позволяет судить о всей группе, и в качестве меры центральной тенденции следовало бы избрать медиану или моду.

Меры центральной тенденции чаще всего используются для сравнения групп по уровню выраженности признака. Если исследователь при этом сомневается, какую меру использовать, то можно дать простые советы.

Выборочные средние можно сравнивать, если выполняются следующие условия:

- группы достаточно большие, чтобы судить о форме распределения;
- распределения симметричны;
- отсутствуют «выбросы».

Если хотя бы одно из перечисленных условий не выполняется, то следует ограничиться модой и медианой. Альтернативой является «сквозное» ранжирование представителей сравниваемых групп и сравнение средних, вычисленных для рангов этих групп.

КВАНТИЛИ РАСПРЕДЕЛЕНИЯ

Помимо мер центральной тенденции в психологии широко используются меры положения, которые называются квантилями распределения. **Квантиль** — это точка на числовой оси измеренного признака, которая делит всю совокупность упорядоченных измерений на две группы с известным соотношением их численности. С одним из квантилей мы уже знакомы — это медиана. Это значение признака, которое делит всю совокупность измерений на две группы с равной численностью. Кроме медианы часто используются процентили и квартили.

Процентили (*Percentiles*) — это 99 точек — значений признака (P_1, ..., P_{99}), которые делят упорядоченное (по возрастанию) множество наблюдений на 100 частей, равных по численности. Определение конкретного значения процентиля аналогично определению медианы. Например, при определении 10-го процентиля, P_{10}, сначала все значения признака упорядочиваются по возрастанию. Затем отсчитывается 10% испытуемых, имеющих наименьшую выраженность признака. P_{10} будет соответствовать тому значению признака, который отделяет эти 10% испытуемых от остальных 90%.

Квартили (*Quartiles*) — это 3 точки — значения признака (P_{25}, P_{50}, P_{75}), которые делят упорядоченное (по возрастанию) множество наблюдений на 4 равные по численности части. Первый квартиль соответствует 25-му процентилю, второй — 50-му процентилю или медиане, третий квартиль соответствует 75-му процентилю.

Процентили и квартили используются для определения частоты встречаемости тех или иных значений (или интервалов) измеренного признака или для выделения подгрупп и отдельных испытуемых, наиболее типичных или нетипичных для данного множества наблюдений.

МЕРЫ ИЗМЕНЧИВОСТИ

Меры центральной тенденции отражают уровень выраженности измеренного признака. Однако не менее важной характеристикой является выраженность индивидуальных различий испытуемых по измеренному признаку. *Меры изменчивости* (*Dispersion*) применяются в психологии для численного выражения величины межиндивидуальной вариации признака.

Наиболее простой и очевидной мерой изменчивости является размах, указывающий на диапазон изменчивости значений. *Размах* (*Range*) — это просто разность максимального и минимального значений:

$$R = x_{max} - x_{min}.$$

Ясно, что это очень неустойчивая мера изменчивости, на которую влияют любые возможные «выбросы». Более устойчивыми являются разновидности размаха: *размах от 10 до 90-го процентиля* ($P_{90} - P_{10}$) или *междуквартильный размах* ($P_{75} - P_{25}$). Последние две меры изменчивости находят свое применение для описания вариации в порядковых данных. А для метрических данных используется дисперсия — величина, название которой в науке является синонимом изменчивости.

Дисперсия (*Variance*) — мера изменчивости для метрических данных, пропорциональная сумме квадратов отклонений измеренных значений от их арифметического среднего:

$$\sum_{i=1}^{N}(x_i - M_x)^2 .$$

Чем больше изменчивость в данных, тем больше отклонения значений от среднего, тем больше величина дисперсии. Величина дисперсии получается при усреднении всех квадратов отклонений:

$$\overline{D}_x = \frac{\sum_{i=1}^{N}(x_i - M_x)^2}{N}. \qquad (4.5)$$

Следует отличать *теоретическую* (генеральную) дисперсию — меру изменчивости бесконечного числа измерений (в генеральной совокупности, популяции в целом) и эмпирическую, или *выборочную*, дисперсию — для реально измеренного множества значений признака. Выборочное значение в статистике используется для оценки дисперсии в генеральной совокупности. Выше указана формула для генеральной (теоретической) дисперсии ($\overline{D}_x$), которая, понятно, не вычисляется. Для вычислений используется формула выборочной (эмпирической) дисперсии (D_x), отличающаяся знаменателем:

$$D_x = \frac{\sum_{i=1}^{N}(x_i - M_x)^2}{N-1}. \qquad (4.6)$$

ПРИМЕР

Вычислим дисперсию признака X для выборки $N = 6$:

№	x_i	$(x_i - M_x)$	$(x_i - M_x)^2$
1	4	4 – 3	1
2	2	2 – 3	1
3	4	4 – 3	1
4	1	1 – 3	4
5	5	5 – 3	4
6	2	2 – 3	1
?	18	0	12

$$M_x = 18/6 = 3;\ D_x = 12/(6-1) = 2{,}4$$

Стандартное отклонение (*Std. deviation*) (сигма, среднеквадратическое отклонение) — положительное значение квадратного корня из дисперсии:

$$\sigma_x = \sqrt{D_x} = \sqrt{\frac{\sum_i (x_i - M_x)^2}{N-1}}. \qquad (4.7)$$

На практике чаще используется именно стандартное отклонение, а не дисперсия. Это связано с тем, что сигма выражает изменчивость в исходных единицах измерения признака, а дисперсия — в квадратах исходных единиц.

Свойства дисперсии:

1. Если значения измеренного признака не отличаются друг от друга (равны между собой) — дисперсия равна нулю. Это соответствует отсутствию изменчивости в данных.

2. Прибавление одного и того же числа к каждому значению переменной не меняет дисперсию:

$$D_{x+c} = D_x, \text{ так как } \sum[(x_i + c) - (M_x + c)]^2 = \sum(x_i - M_x)^2.$$

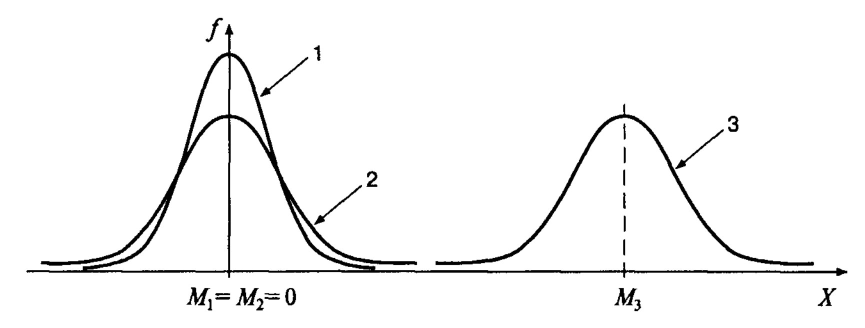


Рис. 4.1. Графики распределения частот: с разной дисперсией ($D_1 < D_2$), одинаковой дисперсией ($D_2 = D_3$) и разными средними арифметическими ($M_2 < M_3$)

Прибавление константы к каждому значению переменной сдвигает график распределения этой переменной на эту константу (меняется среднее), но изменчивость (дисперсия) при этом остается неизменной.

3. Умножение каждого значения переменной на константу c изменяет дисперсию в c^2 раз:

$$D_{x \cdot c} = D_x \cdot c^2, \text{ так как } \sum[(x_i c) - (M_x c)]^2 = c^2 \sum (x_i - M_x)^2.$$

При объединении двух выборок с одинаковой дисперсией, но с разными средними значениями дисперсия увеличивается.

ПРИМЕР

Если одна группа содержит значения: 1, 1, 1, 1, 1, а другая группа — значения 3, 3, 3, 3, 3, то дисперсии этих групп одинаковы и равны 0. Если же объединить эти две группы, то дисперсия будет равна не 0, а 1.

Вообще говоря, справедливо утверждение: при объединении двух групп к внутригрупповой дисперсии каждой группы добавляется дисперсия, обусловленная различием между группами (их средними). И чем больше различие между средними значениями, тем больше увеличивается дисперсия объединенных групп.

Стандартизация или *z-преобразование* данных — это перевод измерений в стандартную *Z-шкалу* (*Z-scores*) со средним $M_z = 0$ и D_z (или σ_z) = 1. Сначала для переменной, измеренной на выборке, вычисляют среднее M_x стандартное отклонение σ_x. Затем все значения переменной x_i пересчитываются по формуле:

$$z_i = \frac{x_i - M_x}{\sigma_x}. \quad (4.8)$$

В результате преобразованные значения (z-значения) непосредственно выражаются в единицах стандартного отклонения от среднего. Если для одной выборки несколько признаков переведены в z-значения, появляется возможность сравнения уровня выраженности разных признаков у того или иного испытуемого. Для того чтобы избавиться от неизбежных отрицательных и дробных значений, можно перейти к любой другой известной шкале: IQ (среднее 100, сигма 15); Т-оценок (среднее 50, сигма 10); 10-балльной — стенов (среднее 5,5, сигма 2) и др. Перевод в новую шкалу осуществляется путем умножения каждого z-значения на заданную сигму и прибавления среднего:

$$S_i = \sigma_S z_i + M_S. \quad (4.9)$$

Асимметрия (***Skewness***) — степень отклонения графика распределения частот от симметричного вида относительно среднего значения. Если исходные данные переведены в z-значения, показатель асимметрии вычисляется по формуле:

$$As = \frac{\sum_i z_i^3}{N}. \quad (4.10)$$

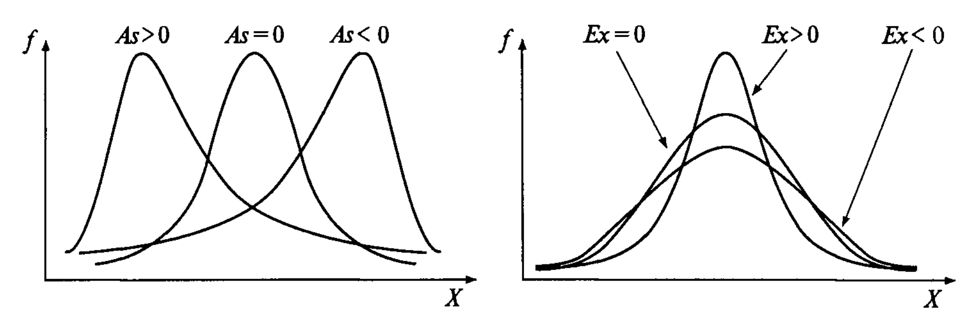


Рис. 4.2. Распределения частот с разными значениями асимметрии и эксцесса

Для симметричного распределения асимметрия равна 0. Если чаще встречаются значения меньше среднего, то говорят о левосторонней, или *положительной* асимметрии ($As > 0$). Если же чаще встречаются значения больше среднего, то асимметрия — правосторонняя, или *отрицательная* ($As < 0$). Чем больше отклонение от нуля, тем больше асимметрия.

Эксцесс (*Kurtosis)* — мера плосковершинности или остроконечности графика распределения измеренного признака. Если исходные данные переведены в z-значения, показатель эксцесса определяется формулой:

$$Ex = \frac{\sum_i z_i^4}{N} - 3\,. \tag{4.11}$$

Островершинное распределение характеризуется положительным эксцессом ($Ex > 0$), а плосковершинное — отрицательным ($-3 < Ex < 0$). «Средневершинное» (нормальное) распределение имеет нулевой эксцесс ($Ex = 0$).

Задачи и упражнения

1. По результатам измерения общительности у юношей (1) и девушек (2) были построены сглаженные графики распределения частот (рис. 4.3).
2. Определите по графику: а) как различаются средние M_1 и M_2; б) как различаются дисперсии D_1 и D_2?
3. Вычислите дисперсии для двух групп:

Группа А	Группа В
3	6
2	5
2	5
1	4

Какой будет дисперсия 8 значений, полученных путем объединения групп? Объясните полученный результат.

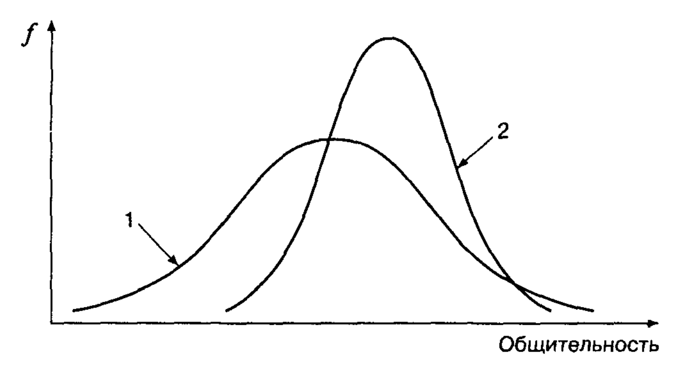


Рис. 4.3. Графики распределения относительных частот общительности юношей (1) и девушек (2)

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Способ 1. Выбираем **Analyze > Descriptive Statistics > Frequencies...** В открывшемся диалоговом окне (**Frequencies**) переносим из левой в правую часть интересующие нас переменные. Если таблица распределения частот нас не интересует, снимаем флажок **Display frequency tables** (Показывать таблицы частот). Нажимаем кнопку **Statistics...** Выбираем интересующие нас статистики и отмечаем их флажком: центральной тенденции (**Central Tendency**) — среднее (**Mean**), моду (**Mode**), медиану (**Median**); изменчивости (**Dispersion**) — стандартное отклонение (**Std. deviation**), дисперсию (**Variance**); распределения — асимметрию (**Skewness**) и эксцесс (**Kurtosis**). После этого нажимаем **Continue**, затем ОК и получаем результат.

Способ 2. Выбираем **Analyze > Descriptive Statistics > Descriptives...** В открывшемся диалоговом окне переносим из левой в правую часть интересующие нас переменные. Нажимаем кнопку **Options...** и отмечаем флажком те статистики, которые нас интересуют (см. выше). Нажимаем **Continue**, затем ОК и получаем результат.

Глава 5

НОРМАЛЬНЫЙ ЗАКОН РАСПРЕДЕЛЕНИЯ И ЕГО ПРИМЕНЕНИЕ

Нормальный закон распределения играет важнейшую роль в применении численных методов в психологии. Он лежит в основе измерений, разработки тестовых шкал, методов проверки гипотез.

История применения закона нормального распределения в социальных и биологических науках начинается, по-видимому, с работы бельгийского ученого А. Кетле «Опыт социальной физики» (1835 г.). В ней он доказывал, что такие явления, как продолжительность жизни, возраст вступления в брак и появления первого ребенка и т. д., подчиняются строгой закономерности. Она проявляется в том, что чаще всего встречаются средние значения соответствующих показателей, и чем больше отклонение от этой средней величины, тем реже встречаемость таких отклонений. Одинаковые отклонения от среднего в меньшую и в большую сторону встречаются одинаково реже, чем среднее значение. Эту закономерность он назвал «законом уклонения от средней величины». В его исследованиях, и позднее — в исследованиях англичанина Ф. Гальтона, было доказано, что распределение частот встречаемости любого демографического (продолжительность жизни и пр.) или антропометрического (рост, вес и пр.) показателя, измеренного на большой выборке людей, имеет одну и ту же «колоколо-

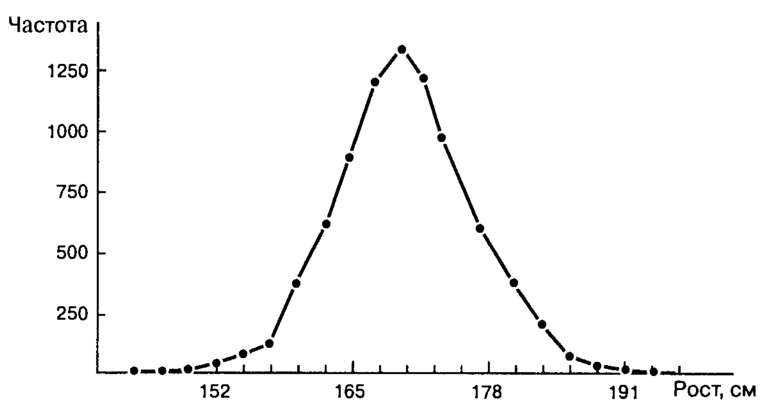


Рис. 5.1. Полигон частот для роста 8585 взрослых людей, родившихся в Англии в XIX в.[1]

[1] *Гласс Дж., Стенли Дж.* Статистические методы в педагогике и психологии. М., 1976. С. 98.



образную» форму (см. рис. 5.1). Форма таких распределений может быть описана математической формулой, которую предложил еще в XVIII веке математик де Муавр.

Де Муавр решал следующую задачу. Предположим, монета в азартной игре подбрасывается 10 раз, и каждый раз она может с равным успехом выпасть «орлом» или «решкой». Какова вероятность того, что в результате этой игры выпадет 0 «орлов», или 1 «орел», ..., 10 «орлов»? Сложные вычисления дают математически точное решение такой задачи (рис. 5.2). А если игра состоит из 100 подбрасываний монеты, или 1000? Де Муавру удалось доказать, что уравнение кривой, соединяющей вершины отрезков на рис. 5.2, для данного случая или для любой другой подобной задачи имеет следующую формулу:

$$f(x_i) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i - M)^2 / 2\sigma^2}, \quad (5.1)$$

где $f(x_i)$ — высота подъема кривой, e — основание натурального логарифма (примерно 2,718), π — число «пи» (примерно 3,14), M и σ — среднее и стандартное отклонения для переменной x_i, которые определяют положение кривой на числовой оси и задают ее размах. Эта формула и соответствующая ей кривая (см. рис. 5.2) впоследствии получили название *закона нормального распределения*.

Итак, исход азартной игры, и продолжительность жизни, и рост человека — все это случайные события, частота (или вероятность) встречаемости которых подчинена закону нормального распределения. А. Кетле объяснял это существованием «идеала» человеческой природы, которому соответствуют средние значения различных пока-



Рис. 5.2. График распределения вероятностей выпадения «орлов» в игре с 10 подбрасываниями монеты и кривая нормального распределения

зателей. Ф. Гальтон, двоюродный брат Ч. Дарвина, проявление нормального закона рассматривал в связи с биологической изменчивостью, наследственностью и отбором. В дальнейшем трудами Ф. Гальтона и его последователей было доказано, что и психологические особенности, например способности, подчиняются нормальному закону. Поэтому дальнейшее развитие измерительного подхода в психологии и статистического аппарата проверки гипотез происходило на базе этого общего закона.

Подведем важный итог этого краткого исторического экскурса. Начиная со второй половины XIX столетия измерительные и вычислительные методы в психологии разрабатываются на основе следующего принципа. *Если индивидуальная изменчивость некоторого свойства есть следствие действия множества причин, то распределение частот для всего многообразия проявлений этого свойства в генеральной совокупности соответствует кривой нормального распределения.* Это и есть закон нормального распределения.

Закон нормального распределения имеет целый ряд очень важных следствий, к которым мы не раз еще будем обращаться. Сейчас же отметим, что если при изучении некоторого свойства мы произвели его измерение на выборке испытуемых и получили отличающееся от нормального распределение, то это значит, что либо выборка нерепрезентативна генеральной совокупности, либо измерения произведены не в шкале равных интервалов.

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ КАК СТАНДАРТ

Каждому психологическому (или шире — биологическому) свойству соответствует свое распределение в генеральной совокупности. Чаще всего оно является нормальным и характеризуется своими параметрами: средним (M) и стандартным отклонением (σ). Только эти два значения отличают друг от друга бесконечное множество нормальных кривых, одинаковой формы, заданной уравнением (5.1). Среднее задает положение кривой на числовой оси и выступает как некоторая исходная, *нормативная величина измерения*. Стандартное отклонение задает ширину этой кривой, зависит от единиц измерения и выступает как *масштаб измерения* (рис. 5.3).

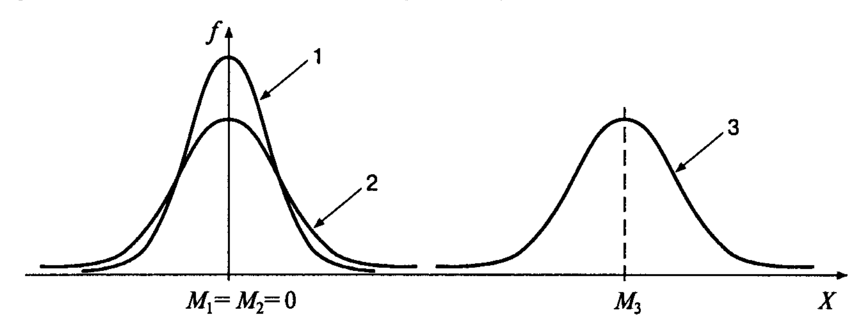


Рис. 5.3. Семейство нормальных кривых, 1-е распределение отличается от 2-го стандартным отклонением ($\sigma_1 < \sigma_2$), 2-е от 3-го средним арифметическим ($M_2 < M_3$)

Все многообразие нормальных распределений может быть сведено к одной кривой, если применить z-преобразование (по формуле 4.8) ко всем возможным измерениям свойств. Тогда каждое свойство будет иметь среднее 0 и стандартное отклонение 1. На рис. 5.4 построен график нормального распределения для $M = 0$ и $\sigma = 1$. Это и есть *единичное нормальное распределение*, которое используется как стандарт — эталон. Рассмотрим его важные свойства.

- Единицей измерения единичного нормального распределения является стандартное отклонение.
- Кривая приближается к оси Z по краям асимптотически — никогда не касаясь ее.
- Кривая симметрична относительно $M = 0$. Ее асимметрия и эксцесс равны нулю.
- Кривая имеет характерный изгиб: точка перегиба лежит точно на расстоянии в одну σ от M.
- Площадь между кривой и осью Z равна 1.

Последнее свойство объясняет название *единичное* нормальное распределение и имеет исключительно важное значение. Благодаря этому свойству *площадь под кривой интерпретируется как вероятность, или относительная частота*. Действительно, вся площадь под кривой соответствует вероятности того, что признак примет любое значение из всего диапазона его изменчивости (от $-\infty$ до $+\infty$). Площадь под единичной нормальной кривой слева или справа от нулевой точки равна 0,5. Это соответствует тому, что половина генеральной совокупности имеет значение признака больше 0, а половина — меньше 0. Относительная частота встречаемости в генеральной совокупности значений признака в диапазоне от z_1 до z_2 равна площади под кривой, лежащей между соответствующими точками. Отметим еще раз, что любое нормальное распределение может быть сведено к единичному нормальному распределению путем z-преобразования.

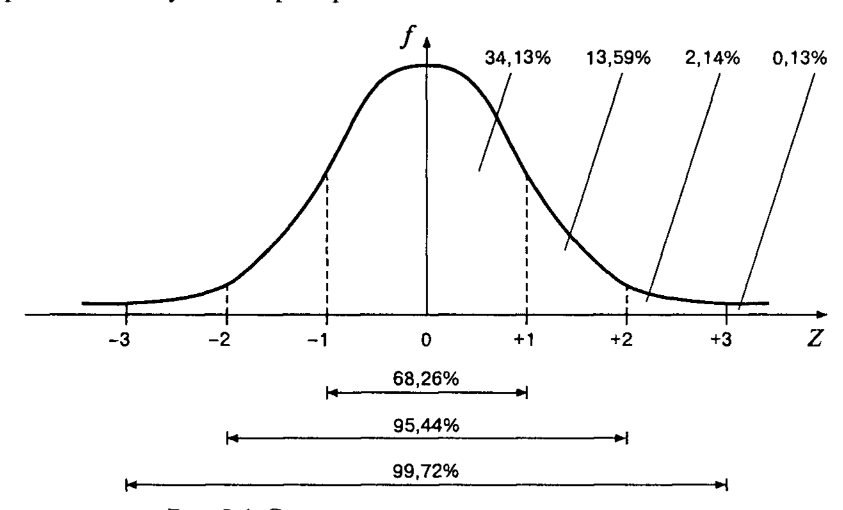
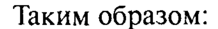


Рис. 5.4. Стандартное нормальное распределение


Таким образом:

- если x_i имеет нормальное распределение со средним M и стандартным отклонением σ, то $z = (x - M_x)/\sigma$ характеризуется единичным нормальным распределением со средним 0 и стандартным отклонением 1;
- площадь между x_1 и x_2 в нормальном распределении со средним M_x и стандартным отклонением σ равна площади между $z_1 = (x_1 - M_x)/\sigma$ и $z_2 = (x_2 - M_x)/\sigma$ в единичном нормальном распределении.

Итак, наиболее важным общим свойством разных кривых нормального распределения является одинаковая доля площади под кривой между одними и теми же двумя значениями признака, выраженными в единицах стандартного отклонения.

Полезно помнить, что для любого нормального распределения существуют следующие соответствия между диапазонами значений и площадью под кривой:

$M \pm \sigma$ соответствует **≈68%** (точно — 68,26%) площади;
$M \pm 2\sigma$ соответствует **≈95%** (точно — 95,44%) площади;
$M \pm 3\sigma$ соответствует **≈100%** (точно — 99,72%) площади.

Единичное нормальное распределение устанавливает четкую взаимосвязь стандартного отклонения и относительного количества случаев в генеральной совокупности для любого нормального распределения. Например, зная свойства единичного нормального распределения, мы можем ответить на следующие вопросы. Какая доля генеральной совокупности имеет выраженность свойства от -1σ до $+1\sigma$? Или какова вероятность того, что случайно выбранный представитель генеральной совокупности будет иметь выраженность свойства, на 3σ превышающую среднее значение? В первом случае ответом будет 68,26% всей генеральной совокупности, так как от –1 до +1 содержится 0,6826 площади единичного нормального распределения. Во втором случае ответ: (100–99,72)/2 = 0,14%.

Полезно знать, что если распределение является нормальным, то:

90% всех случаев располагается в диапазоне значений $M \pm 1{,}64\sigma$;
95% всех случаев располагается в диапазоне значений $M \pm 1{,}96\sigma$;
99% всех случаев располагается в диапазоне значений $M \pm 2{,}58\sigma$.

Существует специальная таблица, позволяющая определять площадь под кривой справа от любого положительного z (приложение 1). Пользуясь ею, можно определить вероятность встречаемости значений признака из любого диапазона. Это широко используется при интерпретации данных тестирования.

ПРИМЕРЫ

1. Значение IQ по шкале Векслера ($M = 100$; $\sigma = 15$) некоторого тестируемого равно 125. Вопрос о *степени выраженности интеллекта* у данного индивидуума переформулируем следующим образом: *насколько часто или редко* встречаются значения IQ ниже или выше 125? Решение. Перейдем от шкалы IQ к единицам

стандартного отклонения (z-значениям): $z = (125 - 100)/15 = 1,66$. По таблице из приложения 1 находим площадь под кривой справа от этого значения, она равна 0,0485. Это значит, что IQ 125 и выше встречается довольно редко — менее, чем в 5% случаев.

2. Какова вероятность того, что случайно выбранный человек будет иметь IQ по шкале Векслера в диапазоне от 100 до 120? Решение. В единицах стандартного отклонения $z_1 = 0,0$; $z_2 = 1,66$. Площадь справа от z_1 – 0,5, справа от z_2 — примерно 0,0918, следовательно, площадь между z_1 и z_2 равна $0,5 - 0,0918 = 0,4082$. Таким образом, вероятность того, что случайно выбранный человек будет иметь IQ в диапазоне от 100 до 120, равна примерно 0,41.

Несмотря на исходный постулат, в соответствии с которым свойства в генеральной совокупности имеют нормальное распределение, реальные данные, полученные на выборке, нечасто распределены нормально. Более того, разработано множество методов, позволяющих анализировать данные без всякого предположения о характере их распределения как в выборке, так и в генеральной совокупности. Эти обстоятельства иногда приводят к ложному убеждению, что нормальное распределение — пустая математическая абстракция, не имеющая отношения к психологии. Тем не менее, как мы увидим в дальнейшем, можно указать по крайней мере на три важных аспекта применения нормального распределения:

1. Разработка тестовых шкал.
2. Проверка нормальности выборочного распределения для принятия решения о том, в какой шкале измерен признак — в метрической или порядковой.
3. Статистическая проверка гипотез, в частности — при определении риска принятия неверного решения.

РАЗРАБОТКА ТЕСТОВЫХ ШКАЛ

Тестовые шкалы разрабатываются для того, чтобы оценить индивидуальный результат тестирования путем сопоставления его с тестовыми нормами, полученными на выборке стандартизации. **Выборка стандартизации** специально формируется для разработки тестовой шкалы — она должна быть репрезентативна генеральной совокупности, для которой планируется применять данный тест. Впоследствии при тестировании предполагается, что и тестируемый, и выборка стандартизации принадлежат одной и той же генеральной совокупности.

Исходным принципом при разработке тестовой шкалы является предположение о том, что измеряемое свойство распределено в генеральной совокупности в соответствии с нормальным законом. Соответственно, измерение в тестовой шкале данного свойства на выборке стандартизации также должно обеспечивать нормальное распределение. Если это так, то тестовая шкала яв-

ляется метрической — точнее, равных интервалов. Если это не так, то свойство удалось отразить в лучшем случае — в шкале порядка. Естественно, что большинство стандартных тестовых шкал являются метрическими, что позволяет более детально интерпретировать результаты тестирования — с учетом свойств нормального распределения — и корректно применять любые методы статистического анализа. *Таким образом, основная проблема стандартизации теста заключается в разработке такой шкалы, в которой распределение тестовых показателей на выборке стандартизации соответствовало бы нормальному распределению.*

Исходные тестовые оценки — это количество ответов на те или иные вопросы теста, время или количество решенных задач и т. д. Они еще называются первичными, или «сырыми» оценками. Итогом стандартизации являются **тестовые нормы** — таблица пересчета «сырых» оценок в стандартные тестовые шкалы.

Существует множество стандартных тестовых шкал, основное назначение которых — представление индивидуальных результатов тестирования в удобном для интерпретации виде. Некоторые из этих шкал представлены на рис. 5.5. Общим для них является соответствие нормальному распределению, а различаются они только двумя показателями: средним значением и масштабом (стандартным отклонением — σ), определяющим дробность шкалы.

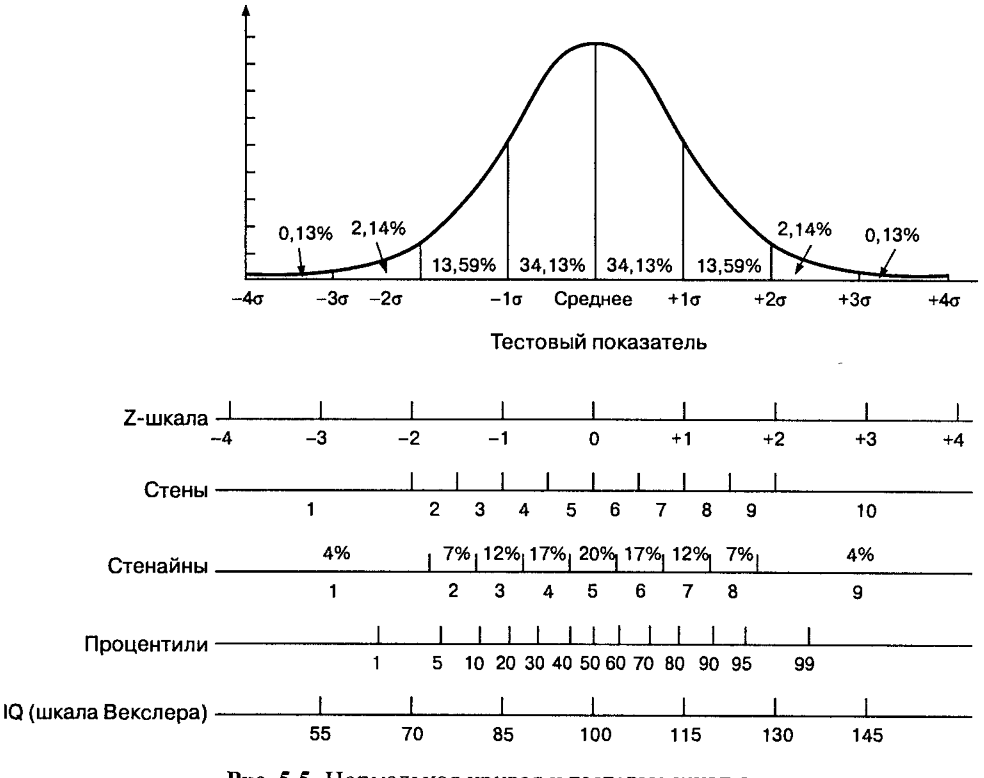


Рис. 5.5. Нормальная кривая и тестовые шкалы

Общая последовательность стандартизации (разработки тестовых норм — таблицы пересчета «сырых» оценок в стандартные тестовые) состоит в следующем:

1) определяется генеральная совокупность, для которой разрабатывается методика и формируется репрезентативная выборка стандартизации;
2) по результатам применения первичного варианта теста строится распределение «сырых» оценок;
3) проверяют соответствие полученного распределения нормальному закону;
4) если распределение «сырых» оценок соответствует нормальному, производится *линейная стандартизация*;
5) если распределение «сырых» оценок не соответствует нормальному, то возможны два варианта:
   - перед линейной стандартизацией производят эмпирическую нормализацию;
   - проводят нелинейную нормализацию.

Проверка распределения «сырых» оценок на соответствие нормальному закону производится при помощи специальных критериев, которые мы рассмотрим далее в этой главе.

Линейная стандартизация заключается в том, что определяются границы интервалов «сырых» оценок, соответствующие стандартным тестовым показателям. Эти границы вычисляются путем прибавления к среднему «сырых» оценок (или вычитания из него) долей стандартных отклонений, соответствующих тестовой шкале. Пример, приведенный ниже, демонстрирует процедуру линейной стандартизации.

ПРИМЕР

Предположим, получено распределение «сырых» оценок, соответствующее нормальному, со средним $M_x = 22$ и стандартным отклонением $\sigma_x = 6$. В качестве стандартной тестовой шкалы выбрана 10-балльная шкала стенов, предложенная Р. Кеттелом ($M_{st} = 5{,}5$; $\sigma_{st} = 2$). Результатом линейной стандартизации должна являться таблица пересчета из шкалы «сырых» оценок в шкалу стенов. Для этого каждому стандартному значению ставится в соответствие интервал «сырых» оценок. Границы интервалов определяются следующим образом. Среднее «сырых» оценок должно делить шкалу стенов ровно пополам (1–5 — ниже среднего, 6–10 — выше среднего). Следовательно, среднее «сырых» оценок $M_x = 22$ — это граница стенов 5 и 6. Следующая граница справа — отделяющая стены 6 и 7 — отстоит от среднего на $\sigma_{st}/2$. Этой границе должна соответствовать граница «сырых» оценок $M_x + \sigma_x/2 = 22 + 3 = 25$. Так же определяются границы всех оставшихся интервалов, а границы крайних интервалов остаются открытыми. Результатом являются тестовые нормы — таблица пересчета «сырых» баллов в стандартные тестовые оценки (табл. 5.1)[1].

---

[1] Обратите внимание, что левая граница каждого диапазона «сырых» оценок исключает границу интервалов, а правая — включает ее. Можно было бы сделать и наоборот, но главное, чтобы границы соседних диапазонов не совпадали, во избежание недоразумений при попадании индивидуального значения на границу интервалов.

Таблица 5.1

Тестовые нормы — таблица пересчета «сырых» баллов в стены

Стены	1	2	3	4	5	6	7	8	9	10
«Сырые» баллы	<11	11–13	14–16	17–19	20–22	23–25	26–28	29–31	32–34	>34

Пользуясь этой таблицей тестовых норм индивидуальный результат («сырой» балл) переводят в шкалу стенов, что позволяет интерпретировать выраженность измеряемого свойства.

В общем случае границы интервалов определяются по формуле z-преобразования:

$$z = \frac{x_i - M_x}{\sigma_x} = \frac{st_i - M_{st}}{\sigma_{st}} \to x_i = M_x + \frac{\sigma_x}{\sigma_{st}}(st_i - M_{st}),$$

где x_i — искомая граница интервала «сырых» оценок, st_i — граница интервала в стандартной тестовой шкале, M_x, σ_x, M_{st}, σ_{st} — средние и стандартные отклонения «сырых» оценок (x) и стандартной шкалы (st).

Эмпирическая нормализация применяется, когда распределение «сырых» баллов отличается от нормального. Она заключается в изменении содержания тестовых заданий. Например, если «сырая» оценка — это количество задач, решенных испытуемыми за отведенное время, и получено распределение с правосторонней асимметрией, то это значит, что слишком большая доля испытуемых решает больше половины заданий. В этом случае необходимо либо добавить более трудные задания, либо сократить время решения.

Нелинейная нормализация применяется, если эмпирическая нормализация невозможна или нежелательна, например, с точки зрения затрат времени и ресурсов. В этом случае перевод «сырых» оценок в стандартные производится через нахождение процентильных границ групп в исходном распределении, соответствующих процентильным границам групп в нормальном распределении стандартной шкалы. Каждому интервалу стандартной шкалы ставится в соответствие такой интервал шкалы «сырых» оценок, который содержит ту же процентную долю выборки стандартизации. Величины долей определяются по площади под единичной нормальной кривой, заключенной между соответствующими данному интервалу стандартной шкалы z-оценками.

Например, для того чтобы определить, какой «сырой» балл должен соответствовать нижней границе стена 10, необходимо сначала выяснить, какому z-значению соответствует эта граница ($z = 2$). Затем по таблице нормального распределения (приложение 1) надо определить, какая доля площади под нормальной кривой находится правее этого значения (0,023). После этого определяется, какое значение отсекает 2,3% наибольших значений «сырых» баллов выборки стандартизации. Найденное значение и будет соответствовать границе 9 и 10 стена.

ПРИМЕР

Рассмотрим пример нелинейной нормализации. Допустим, разрабатываемый тест предполагает решение 20 заданий. Объем выборки стандартизации $N = 200$ человек. Сначала строится таблица распределения частот «сырых» оценок (табл. 5.2).

Таблица 5.2

Таблица распределения частот «сырых» оценок

Оценка	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Частота	2	6	4	6	4	8	6	10	10	12	12	16	24	20	14	14	10	14	8

Исходное распределение заметно отличается от нормального — оно имеет правостороннюю асимметрию (рис. 5.6). В качестве стандартной выберем шкалу стенайнов, для каждой градации которой известны процентные доли (см. рис. 5.5). Исходя из этих процентных долей и таблицы распределения «сырых» оценок строится таблица тестовых норм (табл. 5.3). Сначала отбираются 4% испытуемых, решивших наименьшее количество заданий. У нас 8 испытуемых (4%) решили менее 4 заданий. Это число заданий будет соответствовать 1-му стенайну. Второму стенайну будет соответствовать результат следующих 7% (14) испытуемых: от 4 до 6 заданий, и т. д. Итог нелинейной стандартизации — таблица перевода «сырых» оценок в шкальные, стенайны (табл. 5.3).

Таблица 5.3

Пример нелинейной нормализации: пересчет «сырых» оценок в шкалу стенайнов

Стенайны	1	2	3	4	5	6	7	8	9
%	4	7	12	17	20	17	12	7	4
«Сырые» оценки	<4	4–6	7–9	10–12	13–14	15–16	17–18	19	20

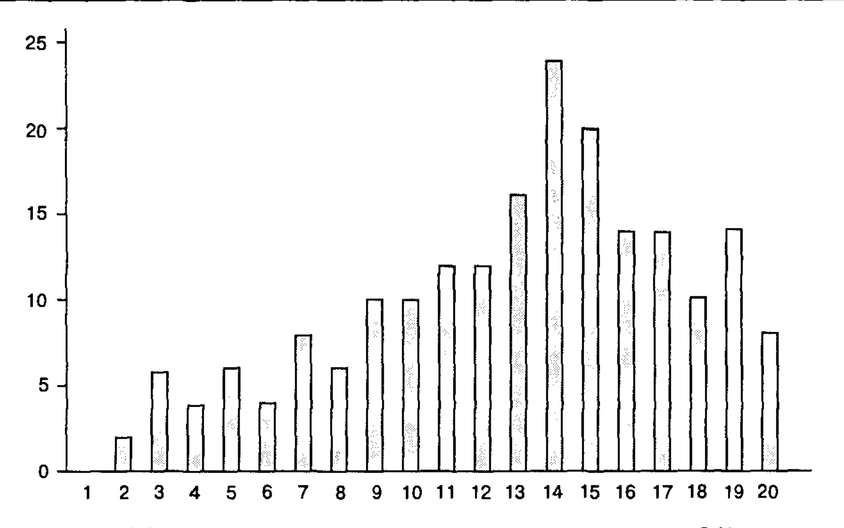


Рис. 5.6. Распределение «сырых» оценок (по данным табл. 5.2)

Изложенные основы психодиагностики позволяют сформулировать математически обоснованные требования к тесту. *Тестовая методика должна содержать:*

- описание выборки стандартизации;
- характеристику распределения «сырых» баллов с указанием среднего и стандартного отклонения;
- наименование, характеристику стандартной шкалы;
- тестовые нормы — таблицы пересчета «сырых» баллов в шкальные.

ПРОВЕРКА НОРМАЛЬНОСТИ РАСПРЕДЕЛЕНИЯ

Для проверки нормальности используются различные процедуры, позволяющие выяснить, отличается ли от нормального выборочное распределение измеренной переменной. Необходимость такого сопоставления возникает, когда мы сомневаемся в том, в какой шкале представлен признак — в порядковой или метрической. А сомнения такие возникают очень часто, так как заранее нам, как правило, не известно, в какой шкале удастся измерить изучаемое свойство (исключая, конечно, случаи явно номинативного измерения).

Важность определения того, в какой шкале измерен признак, трудно переоценить, по крайней мере, по двум причинам. От этого зависит, *во-первых*, полнота учета исходной эмпирической информации (в частности, об индивидуальных различиях), *во-вторых*, доступность многих методов анализа данных. Если исследователь принимает решение об измерении в порядковой шкале, то неизбежное последующее ранжирование ведет к потере части исходной информации о различиях между испытуемыми, изучаемыми группами, о взаимосвязях между признаками и т. д. Кроме того, метрические данные позволяют использовать значительно более широкий набор методов анализа и, как следствие, сделать выводы исследования более глубокими и содержательными.

Наиболее весомым аргументом в пользу того, что признак измерен в метрической шкале, является соответствие выборочного распределения нормальному. Это является следствием закона нормального распределения. *Если выборочное распределение не отличается от нормального, то это значит, что измеряемое свойство удалось отразить в метрической шкале* (обычно — интервальной).

Существует множество различных способов проверки нормальности, из которых мы кратко опишем лишь некоторые, предполагая, что эти проверки читатель будет производить при помощи компьютерных программ.

Графический способ *(Q-Q Plots, P-P Plots)*. Строят либо квантильные графики, либо графики накопленных частот. *Квантильные графики (Q-Q Plots)* строятся следующим образом. Сначала определяются эмпирические значения изучаемого признака, соответствующие 5, 10, ..., 95-процентилю. Затем по таблице нормального распределения для каждого из этих процентилей определяются z-значения (теоретические). Два полученных ряда чисел задают координаты точек на графике: эмпирические значения признака от-

кладываются на оси абсцисс, а соответствующие им теоретические значения — на оси ординат. Для нормального распределения все точки будут лежать на одной прямой или рядом с ней. Чем больше расстояние от точек до прямой линии, тем меньше распределение соответствует нормальному. *Графики накопленных частот (P-P Plots)* строятся подобным образом. На оси абсцисс через равные интервалы откладываются значения накопленных относительных частот, например 0,05; 0,1; ...; 0,95. Далее определяются эмпирические значения изучаемого признака, соответствующие каждому значению накопленной частоты, которые пересчитываются в z-значения. По таблице нормального распределения определяются теоретические накопленные частоты (площадь под кривой) для каждого из вычисленных z-значений, которые откладываются на оси ординат. Если распределение соответствует нормальному, полученные на графике точки лежат на одной прямой.

Критерии асимметрии и эксцесса. Эти критерии определяют допустимую степень отклонения эмпирических значений асимметрии и эксцесса от нулевых значений, соответствующих нормальному распределению. Допустимая степень отклонения — та, которая позволяет считать, что эти статистики существенно не отличаются от нормальных параметров. Величина допустимых отклонений определяется так называемыми стандартными ошибками асимметрии и эксцесса. Для формулы асимметрии (4.10) стандартная ошибка определяются по формуле:

$$As_{sd} = 3\sqrt{\frac{6(N-1)}{(N+1)(N+3)}}\ .$$

Для формулы эксцесса (4.11) стандартная ошибка эксцесса:

$$Ex_{sd} = 5\sqrt{\frac{24N(N-2)(N-3)}{(N+1)^2(N+3)(N+5)}}\ ,$$

где N — объем выборки.

Выборочные значения асимметрии и эксцесса значительно отличаются от нуля, если не превышают значения своих стандартных ошибок. Это можно считать признаком соответствия выборочного распределения нормальному закону. Следует отметить, что компьютерные программы вычисляют показатели асимметрии, эксцесса и соответствующие им стандартные ошибки по другим, более сложным формулам.

Статистический критерий нормальности Колмогорова-Смирнова считается наиболее состоятельным для определения степени соответствия эмпирического распределения нормальному. Он позволяет оценить вероятность того, что данная выборка принадлежит генеральной совокупности с нормальным распределением. Если эта вероятность $p \leqslant 0{,}05$, то данное эмпирическое распределение существенно отличается от нормального, а если $p > 0{,}05$, то делают вывод о приблизительном соответствии данного эмпирического распределения нормальному.

Причины отклонения от нормальности. Общей причиной отклонения формы выборочного распределения признака от нормального вида чаще всего является особенность процедуры измерения: используемая шкала может обладать неравномерной чувствительностью к измеряемому свойству в разных частях диапазона его изменчивости.

ПРИМЕР

Предположим, выраженность некоторой способности определяется количеством выполненных заданий за отведенное время. Если задания простые или время слишком велико, то данная измерительная процедура будет обладать достаточной чувствительностью лишь в отношении части испытуемых, для которых эти задания достаточно трудны. И слишком большая доля испытуемых будет решать все или почти все задания. В итоге мы получим распределение с выраженной правосторонней асимметрией. Можно, конечно, впоследствии повысить качество измерения путем эмпирической нормализации, добавив более сложные задания или сократив время выполнения данного набора заданий. Если же мы чрезмерно усложним измерительную процедуру, то возникнет обратная ситуация, когда большая часть испытуемых будет решать малое количество заданий и эмпирическое распределение приобретет левостороннюю асимметрию.

Таким образом, такие отклонения от нормального вида, как право- или левосторонняя асимметрия или слишком большой эксцесс (больше 0), связаны с относительно низкой чувствительностью измерительной процедуры в области моды (вершины графика распределения частот).

Последствия отклонения от нормальности. Следует отметить, что задача получения эмпирического распределения, строго соответствующего нормальному закону, нечасто встречается в практике исследования. Обычно такие случаи ограничиваются разработкой новой измерительной процедуры или тестовой шкалы, когда применяется эмпирическая или нелинейная нормализация для «исправления» эмпирического распределения. *В большинстве случаев соответствие или несоответствие нормальности является тем свойством измеренного признака, который исследователь должен учитывать при выборе статистических процедур анализа данных.*

В общем случае при значительном отклонении эмпирического распределения от нормального следует отказаться от предположения о том, что признак измерен в метрической шкале. Но остается открытым вопрос о том, какова мера существенности этого отклонения? Кроме того, разные методы анализа данных обладают различной чувствительностью к отклонениям от нормальности. Обычно при обосновании перспективности этой проблемы приводят принцип Р. Фишера, одного из «отцов-основателей» современной статистики: *«Отклонения от нормально-*


Заметно ли "на глаз" отличие распределения от нормального вида?

го вида, если только они не слишком заметны, можно обнаружить лишь для больших выборок; сами по себе они вносят малое отличие в статистические критерии и другие вопросы»[1]. К примеру, при малых, но обычных для психологических исследований выборках (до 50 человек) критерий Колмогорова-Смирнова недостаточно чувствителен при определении даже весьма заметных «на глаз» отклонений от нормальности. В то же время некоторые процедуры анализа метрических данных вполне допускают отклонения от нормального распределения (одни — в большей степени, другие — в меньшей). В дальнейшем при изложении материала мы при необходимости будем оговаривать меру жесткости требования нормальности.

Задачи и упражнения

1. Некоторое свойство измеряется при помощи тестовой шкалы СЕЕВ ($M = 500$, $\sigma = 100$). Какая приблизительно доля генеральной совокупности имеет балл от 600 до 700?
2. В генеральной совокупности значения IQ в шкале Векслера распределены приблизительно нормально со средним 100 и стандартным отклонением 15. С помощью таблиц определите следующие вероятности:
   а) вероятность того, что случайно выбранный человек будет иметь IQ между 79 и 121;
   б) вероятность того, что случайно выбранный человек будет иметь IQ выше 127; ниже 73.
3. Определите при помощи квантильного графика, соответствует ли нормальному виду распределение переменной со следующими значениями процентилей:

Процентили	P_{10}	P_{30}	P_{50}	P_{70}	P_{90}
x_i	6	8	10	11	12

В области каких значений шкала, в которой измерен признак, обладает большей дифференцирующей способностью (чувствительностью), а в какой — меньшей?

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Критерии асимметрии и эксцесса. Выбираем **Analyze > Descriptive Statistics > Descriptives...** В окне диалога переносим из левого окна в правое интересующие нас переменные. Нажимаем кнопку **Options...**, ставим флажок **Distribution >**

[1] Цит. по: Справочник по прикладной статистике: В 2 т. / Под ред. Э. Ллойда, У. Ледермана. М., 1989. Т. 1. С. 270.

Kurtosis, Skewness, нажимаем **Continue**, затем OK. В таблице результатов столбцы **Kurtosis** и **Skewness** содержат значения асимметрии (**Kurtosis**) и эксцесса (**Skewness**) и соответствующие им стандартные ошибки (**Std. Error**). *Распределение соответствует нормальному виду, если для соответствующей переменной абсолютные значения асимметрии и эксцесса не превышают свои стандартные ошибки.*

Графический способ. Выбираем **Graphs > PP...** — графики накопленных частот (или **Graphs > QQ...** — квантильные графики). Открывается диалог **P-P Plots (Q-Q Plots)**. Переносим из левого в правое окно интересующие нас переменные. Нажимаем OK. В окне результатов просматриваем графики **Normal P-P Plot... (Normal Q-Q Plot...)**, на которых по горизонтальной оси отложены соответствующие эмпирические значения, а по вертикальной оси — теоретические значения. *Чем ближе точки графиков к прямой линии, тем меньше отличие распределения от нормального вида.*

Критерий нормальности Колмогорова-Смирнова. Выбираем **Analyze > Nonparametric Tests > 1-Sample K-S...** Открывается диалог **One-Sample Kolmogorov-Smirnov Test**. Переносим из левого в правое окно интересующие нас переменные. Нажимаем OK. В соответствующем переменной столбце находим **Kolmogorov-Smirnov Z** (значение критерия) и **Asymp. Sig. (2-tailed)** (вероятность того, что распределение соответствует нормальному виду). Если значение **Asymp. Sig.** меньше или равно 0,05, то распределение существенно отличается от нормального вида. Если **Asymp. Sig.** больше 0,05, то существенного отличия от нормальности не обнаружено.

Глава 6

КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ

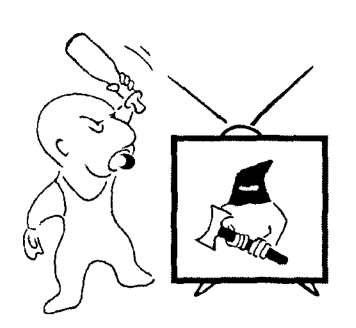
В главе 4 мы рассмотрели основные одномерные описательные статистики — меры центральной тенденции и изменчивости, которые применяются для описания одной переменной. В этой главе мы рассмотрим основные коэффициенты корреляции.

Коэффициент корреляции — двумерная описательная статистика, количественная мера взаимосвязи (совместной изменчивости) двух переменных.

История разработки и применения коэффициентов корреляции для исследования взаимосвязей фактически началась одновременно с возникновением измерительного подхода к исследованию индивидуальных различий — в 1870–1880 гг. Пионером в измерении способностей человека, как и автором самого термина «коэффициент корреляции», был Френсис Гальтон, а самые популярные коэффициенты корреляции были разработаны его последователем Карлом Пирсоном. С тех пор изучение взаимосвязей с использованием коэффициентов корреляции является одним из наиболее популярных в психологии занятием.

К настоящему времени разработано великое множество различных коэффициентов корреляции, проблеме измерения взаимосвязи с их помощью посвящены сотни книг. Поэтому, не претендуя на полноту изложения, мы рассмотрим лишь самые важные, действительно незаменимые в исследованиях меры связи — r-Пирсона, r-Спирмена и τ-Кендалла[1]. *Их общей особенностью является то, что они отражают взаимосвязь двух признаков, измеренных в количественной шкале — ранговой или метрической.*

Вообще говоря, любое эмпирическое исследование сосредоточено на изучении взаимосвязей двух или более переменных.



ПРИМЕРЫ

Приведем два примера исследования влияния демонстрации сцен насилия по ТВ на агрессивность подростков.

1. Изучается взаимосвязь двух переменных, измеренных в количественной (ранговой или метрической) шкале: 1) «время просмотра телепередач с насилием»; 2) «агрессивность».

[1] Читается как тау-Кендалла.

2. Изучается различие в агрессивности 2-х или более групп подростков, отличающихся длительностью просмотра телепередач с демонстрацией сцен насилия.

Во втором примере изучение различий может быть представлено как исследование взаимосвязи 2-х переменных, одна из которых — номинативная (длительность просмотра телепередач). И для этой ситуации также разработаны свои коэффициенты корреляции.

Любое исследование можно свести к изучению корреляций, благо изобретены самые различные коэффициенты корреляции для практически любой исследовательской ситуации. Но в дальнейшем изложении мы будем различать два класса задач:

- *исследование корреляций* — когда две переменные представлены в числовой шкале;
- *исследование различий* — когда хотя бы одна из двух переменных представлена в номинативной шкале.

Такое деление соответствует и логике построения популярных компьютерных статистических программ, в которых в меню *Корреляции* предлагаются три коэффициента (r-Пирсона, r-Спирмена и τ-Кендалла), а для решения других исследовательских задач предлагаются методы сравнения групп.

ПОНЯТИЕ КОРРЕЛЯЦИИ

Взаимосвязи на языке математики обычно описываются при помощи функций, которые графически изображаются в виде линий. На рис. 6.1 изображено несколько графиков функций. Если изменение одной переменной на одну единицу всегда приводит к изменению другой переменной на одну и ту же величину, функция является *линейной* (график ее представляет прямую линию); любая другая связь — *нелинейная*. Если увеличение одной переменной связано с увеличением другой, то связь — *положительная (прямая)*; если увеличение одной переменной связано с уменьшением другой, то связь — *отрицательная (обратная)*. Если направление изменения одной переменной не меняется с возрастанием (убыванием) другой переменной, то такая функция — *монотонная*; в противном случае функцию называют *немонотонной*.

Функциональные связи, подобные изображенным на рис. 6.1, являются идеализациями. Их особенность заключается в том, что одному значению одной переменной соответствует строго определенное значение другой переменной. Например, такова взаимосвязь двух физических переменных — веса и длины тела (линейная положительная). Однако даже в физических экспериментах эмпирическая взаимосвязь будет отличаться от функциональной связи в силу неучтенных или неизвестных причин: колебаний состава материала, погрешностей измерения и пр.

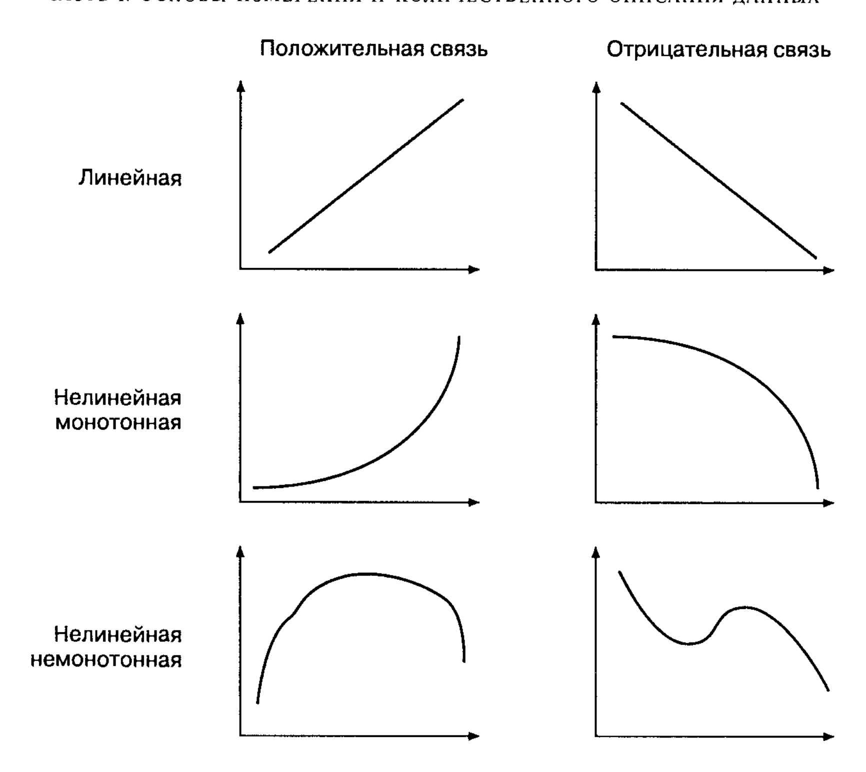


Рис. 6.1. Примеры графиков часто встречающихся функций

В психологии, как и во многих других науках, при изучении взаимосвязи признаков из поля зрения исследователя неизбежно выпадает множество возможных причин изменчивости этих признаков. Результатом является то, что даже *существующая в реальности функциональная связь между переменными выступает эмпирически как вероятностная (стохастическая): одному и тому же значению одной переменной соответствует распределение различных значений другой переменной (и наоборот).* Простейшим примером является соотношение роста и веса людей. Эмпирические результаты исследования этих двух признаков покажут, конечно, положительную их взаимосвязь. Но несложно догадаться, что она будет отличаться от строгой, линейной, положительной — идеальной математической функции, даже при всех ухищрениях исследователя по учету стройности или полноты испытуемых. (Вряд ли на этом основании кому-то придет в голову отрицать факт наличия строгой функциональной связи между длиной и весом тела.)

Итак, в психологии, как и во многих других науках, функциональная взаимосвязь явлений эмпирически может быть выявлена только как вероятностная связь соответствующих признаков. Наглядное представление о характере вероятностной связи дает *диаграмма рассеивания* — график, оси которого соответствуют значениям двух переменных, а каждый испытуемый представляет собой точку (рис. 6.2). *В качестве числовой характеристики вероятностной связи используются коэффициенты корреляции.*

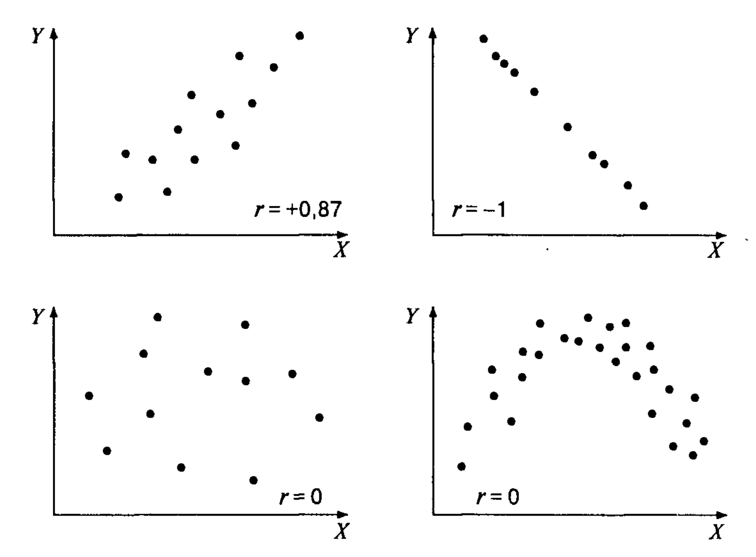


Рис. 6.2. Примеры диаграмм рассеивания и соответствующих коэффициентов корреляции

Коэффициент корреляции — это количественная мера силы и направления вероятностной взаимосвязи двух переменных; принимает значения в диапазоне от –1 до +1.

Сила связи достигает максимума при условии взаимно однозначного соответствия: когда каждому значению одной переменной соответствует только одно значение другой переменной (и наоборот), эмпирическая взаимосвязь при этом совпадает с функциональной линейной связью. Показателем силы связи является *абсолютная* (без учета знака) *величина* коэффициента корреляции.

Направление связи определяется прямым или обратным соотношением значений двух переменных: если возрастанию значений одной переменной соответствует возрастание значений другой переменной, то взаимосвязь называется прямой (положительной); если возрастанию значений одной переменной соответствует убывание значений другой переменной, то взаимосвязь является обратной (отрицательной). Показателем направления связи является *знак* коэффициента корреляции.

КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ *r*-ПИРСОНА

r-Пирсона (Pearson r) применяется для изучения взаимосвязи двух метрических переменных, измеренных на одной и той же выборке. Существует множество ситуаций, в которых уместно его применение. Влияет ли интеллект на успеваемость на старших курсах университета? Связан ли размер заработной платы работника с его доброжелательностью к коллегам? Влияет ли настроение

школьника на успешность решения сложной арифметической задачи? Для ответа на подобные вопросы исследователь должен измерить два интересующих его показателя у каждого члена выборки. Данные для изучения взаимосвязи затем сводятся в таблицу, как в приведенном ниже примере.

ПРИМЕР 6.1

В таблице приведен пример исходных данных измерения двух показателей интеллекта (вербального и невербального) у 20 учащихся 8-го класса.

№	Вербальный IQ (x)	Невербальный IQ (y)
1	13	12
2	9	11
3	8	8
4	9	12
5	7	9
6	9	11
7	8	9
8	13	13
9	11	9
10	12	10
11	8	9
12	9	8
13	10	10
14	10	12
15	12	10
16	10	10
17	8	11
18	9	10
19	10	11
20	11	13
Средние:	9,8	10,4

Связь между этими переменными можно изобразить при помощи диаграммы рассеивания (см. рис. 6.3). Диаграмма показывает, что существует некоторая взаимосвязь измеренных показателей: чем больше значения вербального интеллекта, тем (преимущественно) больше значения невербального интеллекта.

Прежде чем дать формулу коэффициента корреляции, попробуем проследить логику ее возникновения, используя данные примера 6.1. Положение каждой i-точки (испытуемого с номером i) на диаграмме рассеивания относительно остальных точек (рис. 6.3) может быть задано величинами и знаками отклонений соответствующих значений переменных от своих средних величин: $(x_i - M_x)$ и $(y_i - M_y)$. Если знаки этих отклонений совпадают, то это свидетельствует в пользу положительной взаимосвязи (большим значениям

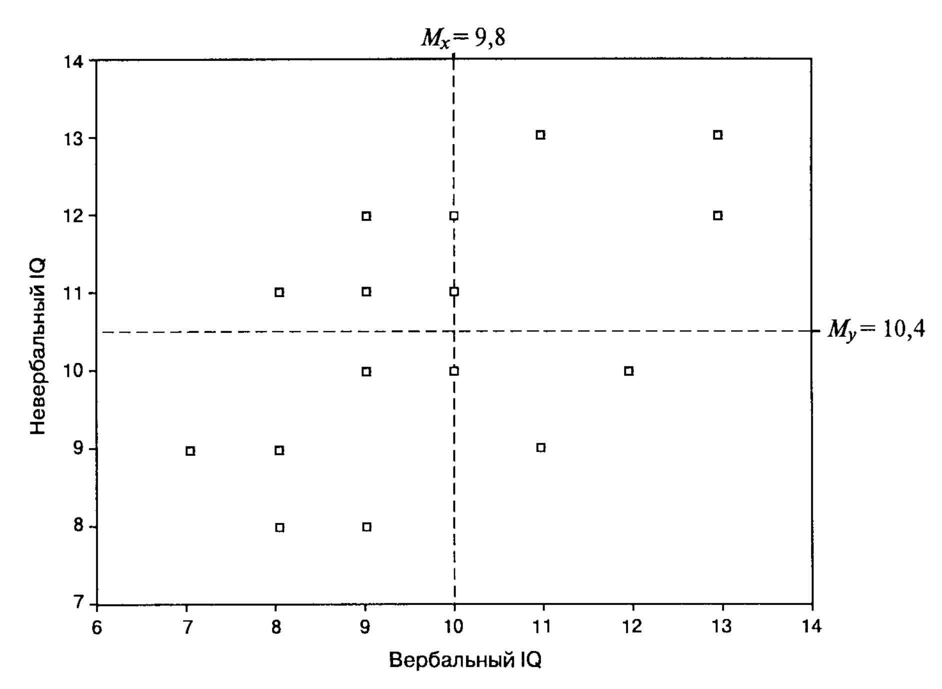


Рис. 6.3. Диаграмма рассеивания для данных примера 6.1

по x соответствуют большие значения по y или меньшим значениям по x соответствуют меньшие значения по y).

ПРИМЕР

Для испытуемого № 1 отклонение от среднего по x и по y положительное, а для испытуемого № 3 и то и другое отклонения отрицательные. Следовательно, данные того и другого свидетельствуют о положительной взаимосвязи изучаемых признаков. Напротив, если знаки отклонений от средних по x и по y различаются, то это будет свидетельствовать об отрицательной взаимосвязи между признаками. Так, для испытуемого № 4 отклонение от среднего по x является отрицательным, по y — положительным, а для испытуемого № 9 — наоборот.

Таким образом, если произведение отклонений $(x_i - M_x) \times (y_i - M_y)$ положительное, то данные i-испытуемого свидетельствуют о прямой (положительной) взаимосвязи, а если отрицательное — то об обратной (отрицательной) взаимосвязи. Соответственно, если x и y в основном связаны прямо пропорционально, то большинство произведений отклонений будет положительным, а если они связаны обратным соотношением, то большинство произведений будет отрицательным. Следовательно, общим показателем для силы и направления взаимосвязи может служить сумма всех произведений отклонений для данной выборки:

$$\sum_{i=1}^{N}(x_i - M_x)(y_i - M_y).$$

При прямо пропорциональной связи между переменными эта величина является большой и положительной — для большинства испытуемых отклонения совпадают по знаку (большим значениям одной переменной соответствуют большие значения другой переменной и наоборот). Если же x и y имеют обратную связь, то для большинства испытуемых большим значениям одной переменной будут соответствовать меньшие значения другой переменной, т. е. знаки произведений будут отрицательными, а сумма произведений в целом будет тоже большой по абсолютной величине, но отрицательной по знаку. Если систематической связи между переменными не будет наблюдаться, то положительные слагаемые (произведения отклонений) уравновесятся отрицательными слагаемыми, и сумма всех произведений отклонений будет близка к нулю.

Чтобы сумма произведений не зависела от объема выборки, достаточно ее усреднить. Но мера взаимосвязи нас интересует не как генеральный параметр, а как вычисляемая его оценка — статистика. Поэтому, как и для формулы дисперсии, в этом случае поступим так же, делим сумму произведений отклонений не на N, а на $N - 1$. Получается мера связи, широко применяемая в физике и технических науках, которая называется *ковариацией (Covariance)*:

$$COV_{xy} = \frac{\sum_{i=1}^{N}(x_i - M_x)(y_i - M_y)}{(N-1)}.$$

В психологии, в отличие от физики, большинство переменных измеряются в произвольных шкалах, так как психологов интересует не абсолютное значение признака, а взаимное расположение испытуемых в группе. К тому же ковариация весьма чувствительна к масштабу шкалы (дисперсии), в которой измерены признаки. Чтобы сделать меру связи независимой от единиц измерения того и другого признака, достаточно разделить ковариацию на соответствующие стандартные отклонения. Таким образом и была получена *формула коэффициента корреляции К. Пирсона*:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - M_x)(y_i - M_y)}{(N-1)\sigma_x\sigma_y}, \qquad (6.1)$$

или, после подстановки выражений для σ_x и σ_y:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - M_x)(y_i - M_y)}{\sqrt{\sum_i (x_i - M_x)^2 \times \sum_i (y_i - M_y)^2}}.$$

Уравнение (6.1) является основной формулой коэффициента корреляции Пирсона. Эта формула вполне осмысленна, но не очень удобна для вычислений «вручную» или на калькуляторе. Поэтому существуют производные фор-

мулы — более громоздкие по виду, менее доступные осмыслению, но упрощающие расчеты. Мы не будем их здесь приводить, так как один раз в жизни можно в учебных целях посчитать корреляцию Пирсона и по исходной формуле «вручную», а в дальнейшем для обработки реальных данных все равно придется воспользоваться компьютерными программами.

ПРИМЕР 6.2

Для расчета коэффициента корреляции воспользуемся данными примера 6.1 о вербальном и невербальном IQ, измеренном у 20 учащихся 8-го класса. К двум столбцам с исходными данными добавляются еще 5 столбцов для дополнительных расчетов, и внизу — строка сумм.

№	X	Y	$(x_i - M_x)$	$(y_i - M_y)$	$(x_i - M_x)^2$	$(y_i - M_y)^2$	$(x_i - M_x)(y_i - M_y)$
1	13	12	3,2	1,6	10,24	2,56	5,12
2	9	11	−0,8	0,6	0,64	0,36	−0,48
3	8	8	−1,8	−2,4	3,24	5,76	4,32
4	9	12	−0,8	1,6	0,64	2,56	−1,28
5	7	9	−2,8	−1,4	7,84	1,96	3,92
6	9	11	−0,8	0,6	0,64	0,36	−0,48
7	8	9	−1,8	−1,4	3,24	1,96	2,52
8	13	13	3,2	2,6	10,24	6,76	8,32
9	11	9	1,2	−1,4	1,44	1,96	−1,68
10	12	10	2,2	−0,4	4,84	0,16	−0,88
11	8	9	−1,8	−1,4	3,24	1,96	2,52
12	9	8	−0,8	−2,4	0,64	5,76	1,92
13	10	10	0,2	−0,4	0,04	0,16	−0,08
14	10	12	0,2	1,6	0,04	2,56	0,32
15	12	10	2,2	−0,4	4,84	0,16	−0,88
16	10	10	0,2	−0,4	0,04	0,16	−0,08
17	8	11	−1,8	0,6	3,24	0,36	−1,08
18	9	10	−0,8	−0,4	0,64	0,16	0,32
19	10	11	0,2	0,6	0,04	0,36	0,12
20	11	13	1,2	2,6	1,44	6,76	3,12
Σ	196	208	0,00	0,00	57,2	42,8	25,6

На первом шаге подсчитываются суммы всех значений одного, затем — другого признака для вычисления соответствующих средних значений M_x и M_y: $M_x = 9{,}8$; $M_y = 10{,}4$.

Далее для каждого испытуемого вычисляются отклонения от среднего: для X и для Y. Каждое отклонение от среднего возводится в квадрат. В последнем столбике записывается результат перемножения двух отклонений от среднего для каждого испытуемого.

Суммы отклонений от среднего для каждой переменной должны быть равны нулю (с точностью до погрешности вычислений). Сумма квадратов отклонений необходима для вычисления стандартных отклонений по известной формуле (4.7):

$$\sigma_x = \sqrt{\frac{57,2}{19}} = 1,735; \; \sigma_y = \sqrt{\frac{42,8}{19}} = 1,501.$$

Сумма произведений отклонений дает нам значение числителя, а произведение стандартных отклонений и ($N-1$) — значение знаменателя формулы коэффициента корреляции:

$$r_{xy} = \frac{25,6}{1,735 \cdot 1,501 \cdot 19} = 0,517.$$

Если значения той и другой переменной были преобразованы в z-значения по формуле:

$$z_{x_i} = \frac{x_i - M_x}{\sigma_x}; \; z_{y_i} = \frac{y_i - M_y}{\sigma_y},$$

то формула коэффициента корреляции r-Пирсона выглядит проще:

$$r_{xy} = \frac{\sum_{i=1}^{N} z_{x_i} z_{y_i}}{N-1}.$$

Отметим еще раз: на величину коэффициента корреляции не влияет то, в каких единицах измерения представлены признаки. Следовательно, *любые линейные преобразования признаков (умножение на константу, прибавление константы: $y_i = x_i b + a$) не меняют значения коэффициента корреляции. Исключением является умножение одного из признаков на отрицательную константу: коэффициент корреляции меняет свой знак на противоположный.*

На рис. 6.2 приведены примеры диаграмм рассеивания для различных значений коэффициента корреляции. Обратите внимание: на последнем рисунке визуально наблюдается нелинейная взаимосвязь между переменными, однако коэффициент корреляции равен нулю. Таким образом, *коэффициент корреляции Пирсона есть мера прямолинейной взаимосвязи; он не чувствителен к криволинейным связям.*

КОРРЕЛЯЦИЯ, РЕГРЕССИЯ И КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ

Корреляция Пирсона есть мера линейной связи между двумя переменными. Она позволяет определить, насколько пропорциональна изменчивость двух переменных. Если переменные пропорциональны друг другу, то графически связь между ними можно представить в виде прямой линии с положительным (прямая пропорция) или отрицательным (обратная пропорция) наклоном. Кроме того, если известна пропорция между переменными, заданная уравнением графика прямой линии:

$$y_i = bx_i + a,$$

то по известным значениям переменной X можно точно предсказать значения переменной Y.

На практике связь между двумя переменными, если она есть, является вероятностной и графически выглядит как облако рассеивания эллипсоидной формы. Этот эллипсоид, однако, можно представить (аппроксимировать) в виде прямой линии, или *линии регрессии*. **Линия регрессии** (*Regression Line*) — это прямая, построенная методом наименьших квадратов: сумма квадратов расстояний (вычисленных по оси Y) от каждой точки графика рассеивания до прямой является минимальной:

$$\sum_i (y_i - \hat{y}_i)^2 = \sum_i e_i^2 = \min,$$

где y_i — истинное i-значение Y, $\hat{y}_i$ — оценка i-значения Y при помощи линии (уравнения) регрессии, $e_i = y_i - \hat{y}_i$ — ошибка оценки (см. рис. 6.4).

Уравнение регрессии имеет вид:

$$\hat{y}_i = bx_i + a, \qquad (6.2)$$

где b — *коэффициент регрессии (Regression Coefficient)*, задающий угол наклона прямой; a — *свободный член*, определяющий точку пересечения прямой оси Y.

Если известны средние, стандартные отклонения и корреляция r_{xy}, то сумма квадратов ошибок минимальна, если:

$$b = r_{xy}\frac{\sigma_y}{\sigma_x}, \quad a = M_y - bM_x. \qquad (6.3)$$

Таким образом, если на некоторой выборке измерены две переменные, которые коррелируют друг с другом, то, вычислив коэффициенты регрессии,

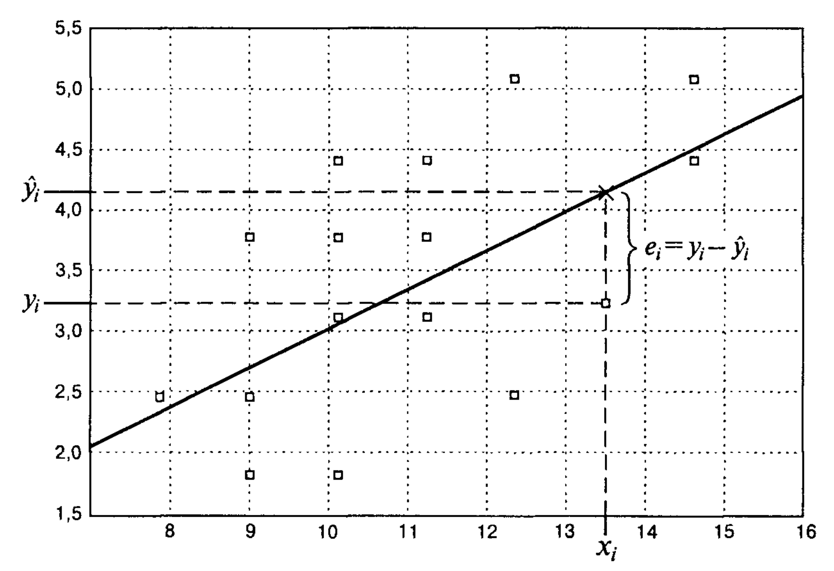


Рис. 6.4. Диаграмма рассеивания и линия регрессии (e_i — ошибка оценки для одного из объектов)

мы получаем принципиальную возможность предсказания неизвестных значений одной переменной (Y — *«зависимая переменная»*) по известным значениям другой переменной (X — *«независимая переменная»*). Например, предсказываемой «зависимой переменной» может быть успешность обучения, а предиктором, «независимой переменной» — результаты вступительного теста.

С какой степенью точности возможно такое предсказание?

Понятно, что наиболее точным предсказание будет, если $|r_{xy}| = 1$. Тогда каждому значению X будет соответствовать только одно значение Y, а все ошибки оценки будут равны 0 (все точки на графике рассеивания будут лежать на прямой регрессии). Если же $r_{xy} = 0$, то $b = 0$ и $\hat{y}_i = M_y$, т. е. при любом X оценка переменной Y будет равна ее среднему значению и предсказательная ценность регрессии ничтожна.

Особое значение для оценки точности предсказания имеет *дисперсия оценок* зависимой переменной. Отметим, что дисперсия оценок равна нулю, если $r_{xy} = 0$ — все оценки равны среднему значению, прямая регрессии параллельна оси X. А если $|r_{xy}| = 1$, то дисперсия оценок равна истинной дисперсии переменной Y, достигая своего максимума:

$$0 \leqslant \sigma^2_{\hat{y}} \leqslant \sigma^2_y .$$

По сути, дисперсия оценок зависимой переменной Y — это та часть ее полной дисперсии, которая обусловлена влиянием независимой переменной X.

Неизвестную дисперсию оценок Y можно выразить через другие, известные статистики, зная рассмотренные ранее свойства дисперсии:

$$\sigma^2_{\hat{y}_i} = \sigma^2_{bx_i+a} = \sigma^2_{bx_i} = b^2\sigma^2_{x_i} ,$$

так как прибавление константы a к каждому значению переменной не меняет дисперсию, а умножение на константу b — увеличивает дисперсию в b^2 раз. Подставляя в формулу выражение для b из (6.2) получаем:

$$\sigma^2_{\hat{y}_i} = r^2_{xy}\frac{\sigma^2_{y_i}}{\sigma^2_{x_i}}\sigma^2_{x_i} = r^2_{xy}\sigma^2_{y_i}, \text{ или}$$

$$r^2_{xy} = \frac{\sigma^2_{\hat{y}_i}}{\sigma^2_{y_i}} . \qquad (6.4)$$

Иначе говоря, *отношение дисперсии оценок зависимой переменной к ее истинной дисперсии равно квадрату коэффициента корреляции.*

Выражение (6.4) дает еще один вариант интерпретации корреляции. Квадрат коэффициента корреляции (R Square) зависимой и независимой переменных представляет долю дисперсии зависимой переменной, обусловленной влиянием независимой переменной, и называется *коэффициентом детерминации*. **Коэффициент детерминации r^2_{xy}**, таким образом, показывает, в какой степени изменчивость одной переменной обусловлена (детерминирована) влиянием другой переменной.

ПРИМЕР

В большинстве исследований взаимосвязи IQ и успеваемости в школе корреляции этих показателей не превышают 0,5–0,7, т. е. коэффициент детерминации достигает величин 0,25–0,49. Иными словами, индивидуальная изменчивость (дисперсия) среднего балла успеваемости может быть предсказана по результатам тестирования IQ не более чем на 25–49%. Означает ли это, что успешность обучения не более чем на 25–49% зависит от интеллекта? Ответ зависит от того, в какой мере средний балл отметок отражает успешность обучения, а тест IQ — интеллектуальные способности учащегося. Во всяком случае, этот пример демонстрирует явно не высокую эффективность двумерной регрессии в деле практического предсказания[1].

Коэффициент детерминации обладает важным преимуществом по сравнению с коэффициентом корреляции. *Корреляция не является линейной функцией связи между двумя переменными.* Поэтому, в частности, среднее арифметическое коэффициентов корреляции для нескольких выборок не совпадает с корреляцией, вычисленной сразу для всех испытуемых из этих выборок (т. е. коэффициент корреляции не аддитивен). Напротив, *коэффициент детерминации отражает связь линейно и поэтому является аддитивным: допускается его усреднение для нескольких выборок.*

Дополнительную информацию о силе связи дает значение коэффициента корреляции в квадрате — коэффициент детерминации r^2: это часть дисперсии одной переменной, которая может быть объяснена влиянием другой переменной. В отличие от коэффициента корреляции r^2 линейно возрастает с увеличением силы связи. На этом основании можно ввести три градации величин корреляции по силе связи:

$r \leqslant 0,3$ — слабая связь (менее 10% от общей доли дисперсии);

$0,3 < r \leqslant 0,7$ — умеренная связь (от 10 до 50% от общей доли дисперсии);

$r > 0,7$ — сильная связь (50% и более от общей доли дисперсии).

ЧАСТНАЯ КОРРЕЛЯЦИЯ

Очень часто две переменные коррелируют друг с другом только за счет того, что обе они согласованно меняются под влиянием некоторой третьей переменной. Иными словами, на самом деле связь между соответствующими свойствами отсутствует, но проявляется в статистической взаимосвязи (корреляции) под влиянием общей причины.

ПРИМЕР

Общей причиной изменчивости двух переменных («третьей переменной») может являться возраст при изучении взаимосвязи различных психологических особенностей в группе детей разного возраста. Предположим, что изучается взаимосвязь между зрелостью моральных суждений — X и скоростью чтения — Y. Но в распоряжении

[1] С более совершенными методами предсказания книга знакомит вас в части 3: «Многомерные методы...»

исследователя имеется лишь выборка из 45 детей разного возраста — от 8 до 14 лет (переменная Z — возраст). Если будет получена существенная положительная корреляция между X и Y, например r_{xy} = 0,54, то о чем это будет свидетельствовать? Осторожный исследователь вряд ли сделает однозначный вывод о том, что зрелость моральных суждений непосредственно связана со скоростью чтения. Скорее всего, дело в том, что и зрелость моральных суждений, и скорость чтения повышаются с возрастом. Иными словами, возраст является причиной согласованной (прямо пропорциональной) изменчивости и зрелости моральных суждений, и скорости чтения.

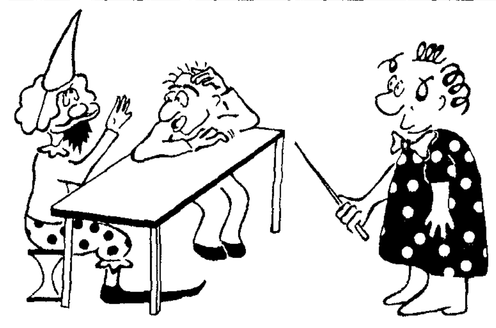
Для численного определения степени взаимосвязи двух переменных при условии исключения влияния третьей применяют **коэффициент частной корреляции** (*Partial Correlation*). Для вычисления частной корреляции достаточно знать три коэффициента корреляции r-Пирсона между переменными X, Y и Z (r_{xy}, r_{xz} и r_{yz}):

$$r_{xy-z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}, \qquad (6.5)$$

где r_{xy-z} — частная корреляция X и Y при постоянном Z (или с учетом Z).

Частная корреляция r_{xy-z} равна r_{xy} при любом фиксированном значении Z (в том случае, если Z линейно коррелирует с X и Y). Например, если значение частной корреляции скорости чтения X и зрелости моральных суждений Y с учетом возраста Z равно 0,2 (r_{xy-z} = 0,2) и возраст линейно коррелирует и с X и с Y, то в любой группе детей одного и того же возраста r_{xy} будет тоже равно 0,2.

ПРИМЕР 6.3



Один исследователь решил сопоставить антропометрические и психологические данные исследования довольно большой группы детей. Каково же было его изумление, когда обнаружилась существенная положительная корреляция между скоростью решения арифметических задач и размером стопы: r_{xy} = 0,42. Оказалось, однако, что дети были разного возраста. Корреляция размера стопы с возрастом составила r_{xy} = 0,7, а корреляция скорости решения арифметических задач с возрастом r_{yz} = 0,6. Эти данные позволяют выяснить, взаимосвязаны ли размер стопы и скорость решения арифметических задач с учетом возраста (при условии, что возраст остается неизменным). Для этого необходимо вычислить частный коэффициент корреляции между размером стопы X и скоростью решения арифметических задач Y (при фиксированном возрасте Z):

$$r_{xy-z} = \frac{0,42 - 0,7 \cdot 0,6}{\sqrt{(1-0,7^2)(1-0,6^2)}} = 0$$

Таким образом, размер стопы и скорость решения арифметических задач коррелируют исключительно за счет согласованности возрастной изменчивости этих показателей: частная корреляция между ними (с учетом возраста) равна нулю. И если мы возьмем группу детей одного и того же возраста, то корреляция размера стопы и скорости решения арифметических задач будет равна нулю.

Следует быть особенно осторожным, пытаясь дать интерпретацию частной корреляции с позиций причинности. Например, если Z коррелирует и с X и с Y, а частная корреляция r_{xy-z} близка к нулю, из этого не обязательно следует, что именно Z является общей причиной для X и Y.

РАНГОВЫЕ КОРРЕЛЯЦИИ

Если обе переменные, между которыми изучается связь, представлены в порядковой шкале, или одна из них — в порядковой, а другая — в метрической, то применяются ранговые коэффициенты корреляции: r-Спирмена или τ-Кенделла. И тот, и другой коэффициент требует для своего применения предварительного ранжирования обеих переменных.

Коэффициент корреляции *r*-Спирмена

Если члены группы численностью N были ранжированы сначала по переменной X, затем — по переменной Y, то корреляцию между переменными X и Y можно получить, просто вычислив коэффициент r-Пирсона для двух рядов рангов. *При условии отсутствия связей в рангах* (т. е. отсутствия повторяющихся рангов) по той и другой переменной, формула для r-Пирсона может быть существенно упрощена в вычислительном отношении и преобразована в формулу, известную как *r-Спирмена*:

$$r_s = 1 - \frac{6\sum_i d_i^2}{N(N^2-1)}, \qquad (6.6)$$

где d_i — разность рангов для испытуемого с номером i.

Коэффициент корреляции r-Спирмена (Spearman's rho) равен коэффициенту корреляции r-Пирсона, вычисленному для двух предварительно ранжированных переменных.

ПРИМЕР 6.4

Предположим, для каждого из 12 учащихся одного класса известно время решения тестовой арифметической задачи в секундах (X) и средний балл отметок по математике за последнюю четверть (Y).

№	X	Y	Ранги X	Ранги Y	d_i	d_i^2
1	122	4,7	7	2	5	25
2	105	4,5	10	4	6	36
3	100	4,4	11	5	6	36
4	145	3,8	5	9	–4	16

№	X	Y	Ранги X	Ранги Y	d_i	d_i^2
5	130	3,7	6	10	–4	16
6	90	4,6	12	3	9	81
7	162	4,0	3	8	–5	25
8	172	4,2	1	6	–5	25
9	120	4,1	8	7	1	1
10	150	3,6	4	11	–7	49
11	170	3,5	2	12	–10	100
12	112	4,8	9	1	8	64
Σ	–	–	78	78	0	474

Для расчета корреляции r-Спирмена сначала необходимо ранжировать учащихся по той и другой переменной. После ранжирования можно проверить его правильность: сумма рангов должна быть равна $N(N+1)/2$. Затем для каждого испытуемого надо вычислить разность рангов (сумма разностей рангов должна быть равна 0). После этого для каждого испытуемого вычисляется квадрат разности рангов — результат приведен в последнем столбце таблицы. Сумма квадратов разностей рангов равна 474. Подставляем известные значения в формулу 6.6:

$$r_s = 1 - \frac{6 \cdot 474}{12(144-1)} = -0{,}657 .$$

Получена умеренная отрицательная связь между успеваемостью по математике и временем решения арифметической задачи.

Отметим: то же значение корреляции было бы получено при использовании формулы r-Пирсона непосредственно к рангам X и Y. Применяя же формулу r-Пирсона к исходным значениям X и Y, мы получим $r_{xy} = -0{,}692$.

Коэффициент корреляции τ-Кендалла

Альтернативу корреляции Спирмена для рангов представляет корреляция τ-*Кендалла*. В основе корреляции, предложенной М. Кендаллом, лежит идея о том, что о направлении связи можно судить, попарно сравнивая между собой испытуемых: если у пары испытуемых изменение по X совпадает по направлению с изменением по Y, то это свидетельствует о положительной связи, если не совпадает — то об отрицательной связи.

В примере 6.3 данные испытуемых 1 и 2 свидетельствуют об отрицательной связи — мы видим *инверсию*: по переменной X у второго испытуемого ранг больше, а по переменной Y — меньше. Данные испытуемых 2 и 3, напротив, демонстрируют *совпадение* направления изменения переменных.

Корреляция τ-Кендалла есть разность относительных частот совпадений и инверсий при переборе всех пар испытуемых в выборке:

$$\tau = P(p) - P(q),$$

где $P(p)$ и $P(q)$ — относительные частоты, соответственно, совпадений и инверсий. Всего в выборке численностью N существует $N(N-1)/2$ всех возможных пар испытуемых. Следовательно,

$$\tau = \frac{P-Q}{N(N-1)/2}, \tag{6.7}$$

где P — число совпадений, Q — число инверсий, а $(P+Q) = N(N-1)/2$.

Формулу 6.7 можно представить и в ином виде:

$$\tau = \frac{P-Q}{P+Q} = 1 - \frac{4Q}{N(N-1)} = \frac{4P}{N(N-1)} - 1. \tag{6.8}$$

При подсчете τ-Кендалла «вручную» данные сначала упорядочиваются по переменной X. Затем для каждого испытуемого подсчитывается, сколько раз его ранг по Y оказывается меньше, чем ранг испытуемых, находящихся ниже. Результат записывается в столбец «Совпадения». Сумма всех значений столбца «Совпадения» и есть P — общее число совпадений, подставляется в формулу 6.8. для вычисления τ-Кендалла.

ПРИМЕР 6.5

Вычислим τ-Кендалла для данных из примера 6.4. Сначала предварительно упорядочиваем испытуемых по переменной X. Затем подсчитываем число совпадений и инверсий для каждого испытуемого, сравнивая по Y его ранг с рангами испытуемых, находящихся под ним. Так, для первого испытуемого ранг по Y равен 6, и 6 испытуемых, находящихся ниже него, имеют по Y более высокий ранг: в столбец «Совпадения» записываем 6. Для третьего по счету испытуемого ранг по Y равен 8, трое испытуемых ниже него имеют более высокий ранг, значит, в столбец «Совпадения» записываем 3, и т. д.

№	Ранги X	Ранги Y	Совпадения	Инверсии
8	1	6	6	5
11	2	12	0	10
7	3	8	3	6
10	4	11	0	8
4	5	9	1	6
5	6	10	0	6
1	7	2	4	1
9	8	7	0	4
12	9	1	3	0
2	10	4	1	1
3	11	5	0	1
6	12	3	0	0
			$P = 18$	$Q = 48$

Проверяем правильность подсчета P и Q: $P + Q = 66$; $N(N-1)/2 = 12 \times 11/2 = 66$.

$$\tau = \frac{18-48}{66} \approx -0{,}455.$$

Для более полной интерпретации полезны соотношения между величиной τ-Кендалла и вероятностью отдельно совпадений и инверсий:

вероятность совпадений $P(p) = \frac{1+\tau}{2}$;

вероятность инверсий $P(q) = 1 - P(p) = \frac{1-\tau}{2}$.

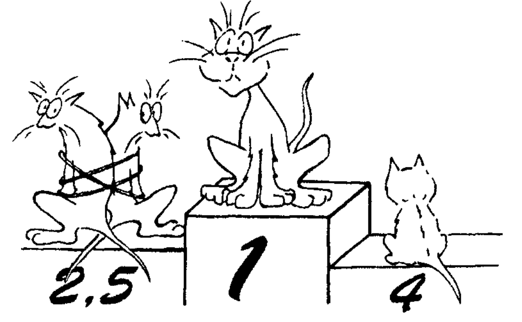
Так, τ = 0,5 значит, что вероятность совпадений равна 0,75, а вероятность инверсий — 0,25, то есть при сравнении объектов друг с другом прямо пропорциональное соотношение (например, роста и веса) встречается в 3 раза чаще, чем обратно пропорциональное соотношение. Такая интерпретация кажется более понятной, чем, например, интерпретация корреляции Пирсона $r = 0{,}5$: «25% изменчивости в весе могут быть объяснены различиями в росте».

τ-Кендалла кажется более простым в вычислительном отношении. Однако при возрастании численности выборки, в отличие от *r*-Спирмена, объем вычислений τ-Кендалла возрастает не пропорционально, а в геометрической прогрессии. Так, при $N = 12$ необходимо перебрать 66 пар испытуемых, а при $N = 48$ — уже 1128 пар, т. е. объем вычислений возрастает более, чем в 17 раз.

Отметим важную особенность ранговых коэффициентов корреляции. Для метрической корреляции *r*-Пирсона значениям +1 или –1 соответствует прямая или обратная пропорция между переменными, что графически представляет собой прямую линию. Максимальным по модулю ранговым корреляциям (+1, –1) вовсе не обязательно соответствуют строгие прямо или обратно пропорциональные связи между исходными переменными *X* и *Y*: достаточна лишь монотонная функциональная связь между ними. Иными словами, ранговые корреляции достигают своего максимального по модулю значения, если большему значению одной переменной всегда соответствует большее значение другой переменной (+1) или большему значению одной переменной всегда соответствует меньшее значение другой переменной и наоборот (–1).

Проблема связанных (одинаковых) рангов

В измерениях часто встречаются одинаковые значения. При их ранжировании возникает проблема *связанных рангов* (*Tied Ranks*). В этом случае действует особое правило ранжирования: объектам с одинаковыми значениями приписывается один и тот же, средний ранг. Например, когда эксперт не может установить различие между двумя лучшими образцами товара, им приписывается одинаковый ранг: (1 + 2)/2 = 1,5. Это сохраняет неизменной сумму рангов для выборки объемом *N*: $N(N+1)/2$.



При наличии одинаковых (связанных) рангов формулы ранговой корре-

ляции Спирмена (6.6) и Кендалла (6.7 и 6.8) не подходят. Хотя сумма рангов и не меняется, но изменчивость данных становится меньше. Соответственно, уменьшается возможность оценить степень связи между измеренными свойствами.

При использовании корреляции Спирмена в случае связанных рангов возможны два подхода:

- если связей немного (менее 10% для каждой переменной), то вычислить *r*-Спирмена приближенно по формуле 6.6;
- при большем количестве связей применить к ранжированным данным классическую формулу *r*-Пирсона 6.1 — это всегда позволит определить ранговую корреляцию независимо от наличия связей в рангах.

При использовании корреляции τ-Кендалла в случае наличия связанных рангов в формулу вносятся поправки, и тогда получается общая формула для вычисления τ коэффициента корреляции τ_b-*Кендалла* (*Kendall's tau-b*) независимо от наличия или отсутствия связей в рангах:

$$\tau_b = \frac{P-Q}{\sqrt{[N(N-1)/2]-K_x}\sqrt{[N(N-1)/2]-K_y}}, \tag{6.9}$$

где $K_x = (1/2)\sum_i f_i(f_i - 1)$ (i — количество групп связей по X, f_i — численность каждой группы); $K_y = (1/2)\sum_i f_i(f_i - 1)$ (i — количество групп связей по Y, f_i — численность каждой группы).

ПРИМЕР 6.6

Супруги X и Y ранжировали 8 жизненных ценностей по степени предпочтения. Данные представлены в таблице:

Ценности	Ранги X	Ранги Y	P (совпадения)	Q (инверсии)
Здоровье	1	1	7	0
Любовь	2	3	4	0
Богатство	4	3	3	0
Свобода	4	3	3	0
Мудрость	4	5	3	0
Познание	6	7	0	0
Развитие	7	7	0	0
Творчество	8	7	0	0
			$\Sigma = 20$	

В качестве меры согласованности предпочтений супругов вычислим корреляцию τ_b-Кендалла, так как наблюдаются связи в рангах: одна группа из трех рангов по X и две группы по три ранга по Y.

Обратите внимание на подсчет совпадений для объектов, попадающих в «связки». Например, для объекта «Богатство» пропускаются два ниже находящихся объекта, как имеющие одинаковые с ним ранги по X.

Подсчитаем поправочные коэффициенты: $K_x = (1/2)[(3(3-1)] = 3$; $K_y = (1/2)[3(3-1) + 3(3-1)] = 6$. Подставим полученные значения в формулу 6.9:

$$\tau_b = \frac{20-0}{\sqrt{(8\times 7/2)-3}\ \sqrt{(8\times 7/2)-6}} = 0{,}853\,.$$

Заметим, что инверсии отсутствуют, и если бы связей в рангах не было, то корреляция была бы строго прямой (равна 1).

КОРРЕЛЯЦИЯ БИНАРНЫХ ДАННЫХ

Как отмечалось ранее, если одна из двух переменных представлена в номинативной шкале, а другая — в числовой (ранговой или метрической), то связь между этими переменными лучше изучать путем сравнения групп по уровню выраженности числовой переменной.

ПРИМЕР

Предположим, исследуется связь количества пропущенных лекций студентами и курса обучения (с 1-го по 5-й). Первая переменная — метрическая, а вторая — номинативная. Связь между этими переменными может быть изучена путем сравнения разных курсов по количеству пропущенных лекций (по средним значениям). Если будут обнаружены различия между курсами, то посещаемость лекций связана с курсом обучения, в противном случае — связи нет.

То же касается проблемы изучения связи между двумя номинативными переменными. Хотя и для этого случая существуют коэффициенты корреляции (*K* — Чупрова, *C* — Пирсона), но возможность их интерпретации весьма ограничена, в частности потому, что они отражают лишь силу связи, но не ее направление. Поэтому и в этом случае проблему связи между двумя номинативными переменными лучше изучать путем сравнения градаций одной переменной по распределению другой переменной.

ПРИМЕР

Предположим, исследуется связь агрессивности учащихся (три градации: низкая, средняя, высокая) и образования их родителей (среднее, высшее техническое, высшее гуманитарное). Результаты исследования связей двух номинативных переменных обычно представляются в виде таблицы сопряженности:

Агрессивность	Образование родителей		
	Среднее	Высш. технич.	Высш. гуманит.
Низкая	15	10	12
Средняя	18	15	14
Высокая	10	8	7

Связь между этими переменными может быть изучена путем сравнения распределений учащихся по степени агрессивности для разных градаций образования родителей (или, что то же самое, путем сравнения распределения образования родителей для разных градаций степени агрессивности учащихся).

Исключением можно считать случай изучения связи двух бинарных переменных. *Бинарная переменная* имеет только две градации, обычно обозначаемые как 0 и 1. Примеры таких переменных: пол (мужской, женский), образование (среднее, высшее), тревожность (низкая, высокая), успешность (низкая, высокая) и т. д.

При изучении связей между бинарными переменными обычно строят четырехклеточные таблицы сопряженности:

Таблица 6.1

Таблица сопряженности 2×2

		Признак X		Итог
		0	1	
Признак Y	0	a	b	$a+b$
	1	c	d	$c+d$
Итог		$a+c$	$b+d$	N

В этом случае допустимо применение r-Пирсона (формула 6.1) непосредственно к исходным данным — двум бинарным переменным, принимающим значение 0 или 1, измеренным для каждого члена выборки численностью N. *Результат применения r-Пирсона к двум бинарным переменным называется «фи-коэффициентом сопряженности»* (*Phi*). Если данные представлены в четырехклеточной таблице сопряженности, то применяется формула, существенно упрощающая расчеты, но дающая аналогичный результат:

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, \quad (6.10)$$

где a, b, c, d соответствуют обозначениям в четырехклеточной таблице 6.1.

ПРИМЕР 6.7

Исследовалась связь семейного положения студенток (X: 0 — холостая, 1 — замужем) и их академической успеваемости (Y: 0 — закончила вуз, 1 — отчислена). В распоряжении исследователя есть данные для 12 студенток:

№	1	2	3	4	5	6	7	8	9	10	11	12
X	0	1	0	0	1	1	0	1	0	0	1	0
Y	0	1	1	0	1	0	0	1	1	0	1	0

Таблица сопряженности для этих данных:

		X		Итог
		0	1	
Y	0	5	1	6
	1	2	4	6
Итог		7	5	12

Вычислим φ-коэффициент сопряженности:

$$\varphi = \frac{5 \cdot 4 - 1 \cdot 2}{\sqrt{6 \cdot 6 \cdot 7 \cdot 5}} = 0{,}507.$$

Получена умеренная положительная взаимосвязь: холостые студентки чаще заканчивают вуз, а замужние — чаще отчисляются. Отметим, что тот же самый результат был бы получен при применении формулы r-Пирсона непосредственно к исходным данным.

Итак, φ-коэффициент есть просто r-Пирсона, вычисленный для бинарных данных, а формула 6.10 алгебраически эквивалентна формуле 6.1. Следовательно, интерпретация φ-коэффициента подобна интерпретации r-Пирсона. Но использование φ-коэффициента существенно ограничено. Чем больше асимметрия распределения 0 и 1 по каждой переменной, тем менее точно φ-коэффициент отражает связь между бинарными переменными. Иначе говоря, *применение φ-коэффициента требует приблизительного равенства количества 0 и 1 по каждой переменной.*

ВЕЛИЧИНА КОРРЕЛЯЦИИ И СИЛА СВЯЗИ

Коэффициенты корреляции были специально разработаны для численного определения силы и направления связи между двумя свойствами, измеренными в числовых шкалах (метрических или ранговых). Как уже упоминалось, максимальной силе связи соответствуют значения корреляции +1 (строгая прямая или прямо пропорциональная связь) и –1 (строгая обратная или обратно пропорциональная связь), отсутствию связи соответствует корреляция, равная нулю. Дополнительную информацию о силе связи дает значение коэффициента детерминации r^2: это часть дисперсии одной переменной, которая может быть объяснена влиянием другой переменной.

Однако в ряде случаев разные коэффициенты корреляции имеют различную эффективность, а иногда все они оказываются нечувствительными к связям.

Выбросы и отклонения распределений от нормальности



Выбросы — это экстремально большие или малые значения признака. В наиболее существенной степени выбросы влияют на корреляцию r-Пирсона, так как величина этого коэффициента прямо пропорциональна отклонению значения переменной от среднего.

ПРИМЕР 6.8

Воспользуемся данными из примера 6.1 с показателями вербального и невербального интеллекта, измеренного у 20 учащихся 8-го класса ($r = 0,517$). Добавим еще одно наблюдение: $x_{21} = 3$, $y_{21} = 16$ (см. рис. 6.5). Новое значение r-Пирсона для всех $N = 21$ теперь будет равно $r = -0,124$.

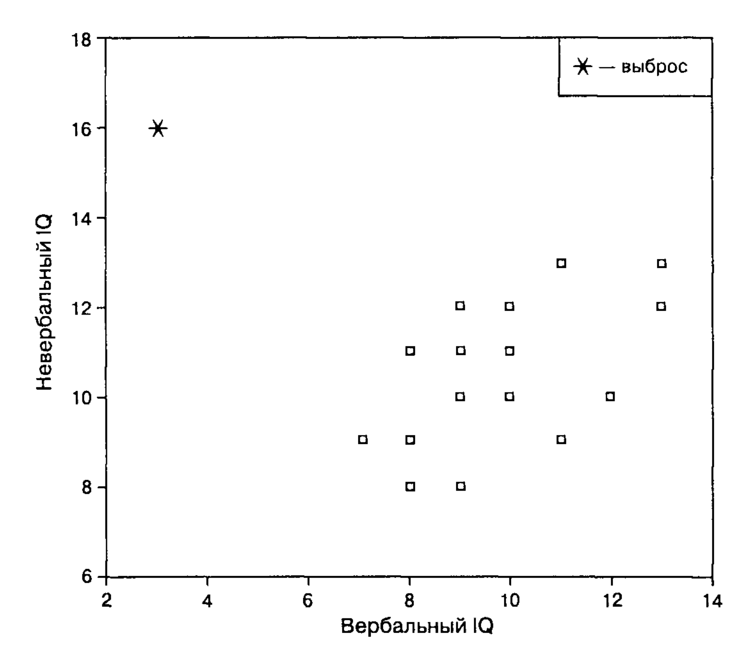


Рис. 6.5. Демонстрация влияния экстремальных значений признаков («выброса») на коэффициент корреляции Пирсона

Пример 6.8 демонстрирует, что даже одно наблюдение с экстремально большими или малыми значениями переменных может изменить знак корреляции на противоположный. Точно так же немногочисленные выбросы могут обусловить и появление корреляции.

Существенно меньшему влиянию выбросов подвержены ранговые корреляции. Поэтому один из способов борьбы с выбросами — переход к рангам и применение ранговых коэффициентов корреляции.

Для примера 6.8 ранговые коэффициенты корреляции (Спирмена и Кендалла) для первых 20 испытуемых (без выброса) составляют, соответственно: $r_s = 0,505$; $\tau = 0,390$. При добавлении выбросов: $r_s = 0,294$; $\tau = 0,239$. Значения корреляций уменьшилось, но не столь существенно, как r-Пирсона.

Другой подход к выбросам подразумевает «чистку» данных. Можно для каждой переменной установить определенное ограничение на диапазон ее изменчивости. Например, исключать те наблюдения, которые выходят за пределы диапазона $M \pm 2\sigma$ (или даже $M \pm 1,5\sigma$).

Часто такая «чистка» совершенно необходима. Например, при исследовании времени реакции, когда основная масса наблюдений находится в диапазоне 250–700 мс, исключение нескольких «странных» значений меньше 50 мс и больше 1000 мс может существенно изменить общую картину.

По сути, наличие выбросов означает отклонение распределений одной или обеих переменных от нормального вида. В общем случае, если распределения переменных сильно скошены (асимметричны), это может существенно снижать значение корреляции даже при сильной связи между соответствующими свойствами или, наоборот, обусловить появление «ложной» корреляции. Особенно сильно асимметричность распределений влияет на r-Пирсона. Поэтому при существенном отклонении формы распределения хотя бы одной переменной от нормального вида желательно перейти к рангам и воспользоваться ранговым коэффициентом корреляции.

Влияние «третьей» переменной

Иногда корреляция между двумя переменными обусловлена не связью между соответствующими свойствами, а влиянием некоторой общей причины совместной изменчивости этих переменных, которая зачастую выпадает из поля зрения исследователя. Эта общая причина может быть измерена как некоторая «третья» переменная, представленная либо в номинативной шкале, либо в количественной (ранговой или метрической) шкале.

Если истинная причина корреляции представляет собой номинативную переменную, то это проявляется в характерной неоднородности выборки: в ней можно обнаружить различные группы, для которых согласованно меняются средние двух переменных, в то время как внутри групп эти переменные не коррелируют. Если подобное явление возможно и существует способ *содержательно интерпретируемого* деления выборки на группы, необходимо вычислить корреляцию не только для всей выборки, но и для каждой группы в отдельности.

ПРИМЕР

Если мы возьмем достаточно большую группу людей — мужчин и женщин, то обнаружим существенную отрицательную корреляцию роста и длины волос: чем больше рост, тем короче волосы. Однако, рассматривая график рассеивания роста и длины волос с выделением групп мужчин и женщин, мы обнаружим истинную причину этой корреляции — пол (рис. 6.6). Корреляции роста и длины волос отдельно для мужчин и отдельно для женщин будут близки к нулю.

Другой случай «ложной» корреляции — когда «третья» переменная может быть представлена в числовой шкале.

ПРИМЕР

Число церквей и количество увеселительных заведений в городах, как известно, сильно коррелируют, так же, впрочем, как рост и навык чтения у детей. Нетрудно

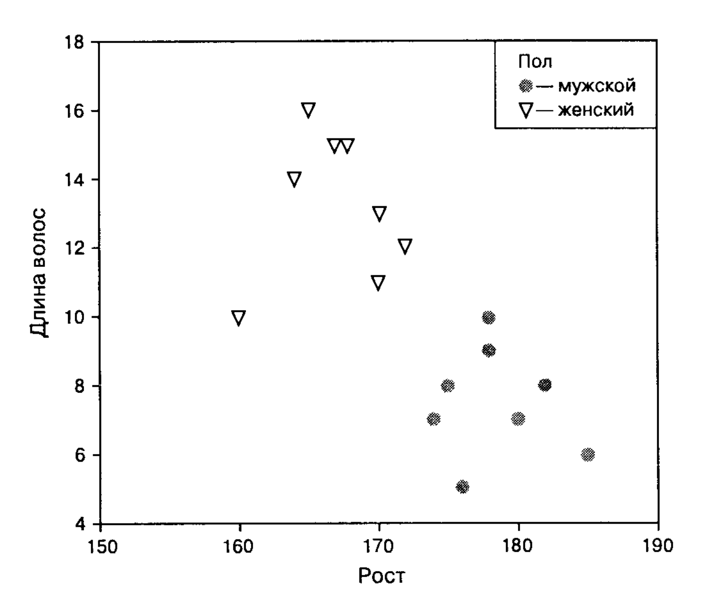


Рис. 6.6. График рассеивания для роста и длины волос. Темные точки — мужчины, светлые треугольники — женщины

догадаться, что в первом случае «третьей» переменной является численность городского населения, а во втором — возраст детей. (См. также пример 6.3 из раздела «Частная корреляция».)

Если истинная причина корреляции между двумя переменными X и Y измерена как количественная переменная Z, то предположение о том, что именно она является причиной корреляции, можно проверить, вычислив частную корреляцию r_{xy-z} по формуле 6.5. Если частная корреляция X и Y с учетом Z (r_{xy-z}) существенно меньше r_{xy}, то весьма вероятно, что именно Z является истинной причиной корреляции X и Y.

Следует отметить, что за редким исключением факт наличия или отсутствия корреляции может быть объяснен влиянием некоторой «третьей» переменной, упущенной из поля зрения исследователя. Таким образом, всегда остается возможность альтернативной интерпретации обнаруженной корреляции.

Нелинейные связи

Еще одним источником низкой эффективности корреляций являются возможный нелинейный характер связи между переменными. То, какой характер имеет связь между переменными, можно заметить, рассматривая график двумерного рассеивания. Это свидетельствует о важности визуального анализа связи с помощью таких графиков во всех случаях применения корреляций.

К отклонениям от прямолинейной зависимости любого рода наиболее чувствителен коэффициент корреляции r-Пирсона. Однако если нелинейная

связь оказывается монотонной, то возможен переход к рангам и применение ранговых корреляций.

Довольно часто в исследованиях встречаются немонотонные связи — когда связь меняет свое направление (с прямого на обратное, или наоборот) при увеличении или уменьшении значений одной из переменной.

ПРИМЕРЫ

Наиболее типичный пример — это связь тревожности и результатов тестирования, или в общем случае — связь уровня активации (X) и продуктивности деятельности (Y). Связь таких переменных напоминает перевернутую (инвертированную) U (рис. 6.7). Любой из рассмотренных коэффициентов корреляции будет в этом случае иметь значение, близкое к нулю.



Рис. 6.7. Пример криволинейной немонотонной связи между уровнем активации и продуктивностью деятельности

Если наблюдается немонотонная нелинейность связи, то можно поступить двояко. В первом случае сначала надо найти точку перегиба по графику рассеивания и разделить выборку на две группы, различающиеся направлением связи между переменными. После этого можно вычислять корреляции отдельно для каждой группы. Второй способ предполагает отказ от применения коэффициентов корреляции. Необходимо ввести дополнительную номинативную переменную, которая делит исследуемую выборку на контрастные группы по одной из переменных. Далее можно изучать различия между этими группами по уровню выраженности (например, по средним значениям) другой переменной.

В приведенном примере (рис. 6.7) можно по переменной «активация» выделить 3 группы (низкий, средний и высокий уровень) и далее изучать различия между этими группами по продуктивности деятельности.

КАКОЙ КОЭФФИЦИЕНТ КОРРЕЛЯЦИИ ВЫБРАТЬ

При изучении связей между переменными наиболее предпочтительным является случай применения *r-Пирсона* непосредственно к исходным данным. В любом случае, обнаружена корреляция или нет, необходим визуальный ана-

лиз графиков распределения переменных и графика двумерного рассеивания, если исследователя действительно интересует связь между соответствующими переменными. Применяя *r*-Пирсона, необходимо убедиться, что:

- обе переменные не имеют выраженной асимметрии;
- отсутствуют выбросы;
- связь между переменными прямолинейная.

Если хотя бы одно из условий не выполняется, можно попытаться применить ранговые коэффициенты корреляции: *r-Спирмена* или *τ-Кендалла*. Но и ранговые корреляции имеют свои ограничения. Они применимы, если:

- обе переменные представлены в количественной шкале (метрической или ранговой);
- связь между переменными является монотонной (не меняет свой знак с изменением величины одной из переменных).

Применение ранговых коэффициентов корреляции при расчете «вручную» требует предварительного ранжирования переменных. Если при этом встречаются одинаковые значения признаков (связи в рангах), применяется формула *r*-Пирсона для предварительно ранжированных переменных (в случае с *r*-Спирмена) либо вводятся поправки на связанные ранги (в случае с τ-Кендалла).

Если есть предположение, что корреляция обусловлена влиянием третьей переменной, и все три переменные допускают применение *r*-Пирсона для вычисления корреляции между ними, возможна проверка этого предположения путем вычисления *коэффициента частной корреляции* этих переменных (при фиксированных значениях третьей переменной). Если значение частной корреляции двух переменных по абсолютной величине заметно меньше, чем их парная корреляция, то парная корреляция обусловлена влиянием третьей переменной.

Применяя коэффициенты корреляции, особое внимание следует уделять *графикам двумерного рассеивания*. Они позволяют выявить случаи, когда корреляция обусловлена неоднородностью выборки по той и другой переменной. Кроме того, эти графики позволяют определить характер связи: ее линейность и монотонность. Если связь является криволинейной и не монотонной (например, имеет форму *U*), то коэффициенты корреляции не подходят. В этом случае можно разделить выборку на группы по одной из переменных, для сравнения этих групп по выраженности другой переменной.

Если обе переменные представлены в бинарной шкале (0,1), для изучения связи между ними можно применять φ-коэффициент сопряженности, если для каждой переменной количество 0 и 1 приблизительно одинаковое.

Во всех случаях, когда исследователя интересует связь между переменными, а коэффициенты корреляции для этого не подходят, изучение этой связи возможно при помощи *сравнения групп*, выделяемых по одной из переменных. Если другая переменная метрическая или ранговая, то группы сравниваются по уровню ее выраженности, если номинативная — то по ее распределению.

ОБРАБОТКА НА КОМПЬЮТЕРЕ

1. Графики двумерного рассеивания. Выбираем **Graphs... > Scatter... > Simple**. Нажимаем **Define**. В появляющемся окне назначаем осям переменные: выделяем слева одну переменную, нажимаем ▹ напротив «**X Axis**» (Ось X), выделяем другую переменную, нажимаем ▹ напротив «**Y Axis**». Нажимаем OK. Получаем график рассеивания назначенных переменных.

2. Вычисление парных корреляций. Выбираем **Analyze > Correlate > Bivariate...** В открывшемся окне диалога переносим интересующие переменные из левой части в правую при помощи кнопки ▹ (переменных должно быть как минимум две). По умолчанию стоит флажок «**Pearson**» (корреляция *r*-Пирсона). Если интересует корреляция *r*-Спирмена или τ-Кендалла, необходимо поставить соответствующие флажки внизу. Нажимаем OK. В появившейся таблице строки и столбцы соответствуют выделенным ранее переменным. В ячейке на пересечении строки и столбца, соответствующих интересующим нас переменным, видим три числа: верхнее соответствует коэффициенту корреляции, нижнее — численности выборки *N*, среднее — уровню значимости.

3. Вычисление частной корреляции. Выбираем **Analyze > Correlate > Partial...** В открывшемся окне диалога переносим интересующие переменные из левой части в правое верхнее окно (**Variables:**) при помощи верхней кнопки ▹ (переменных должно быть как минимум две). Затем при помощи нижней кнопки ▹ из левой части в правое нижнее окно (**Controlling for:**) переносим переменную, значения которой хотим фиксировать. Нажимаем OK. Получаем таблицу, аналогичную таблице парных корреляций, но верхнее число в каждой ячейке — значение частной корреляции соответствующих двух переменных при фиксированном значении указанной третьей переменной. Нижнее число — уровень значимости, а посередине — число степеней свободы.

Часть II

МЕТОДЫ СТАТИСТИЧЕСКОГО ВЫВОДА: ПРОВЕРКА ГИПОТЕЗ

Глава 7

ВВЕДЕНИЕ В ПРОБЛЕМУ СТАТИСТИЧЕСКОГО ВЫВОДА

ГИПОТЕЗЫ НАУЧНЫЕ И СТАТИСТИЧЕСКИЕ

Обычно исследование проводится для проверки гипотезы, которая является следствием теоретических представлений.[1] Эта гипотеза содержит утверждение о связи абстрактных категорий, относящихся к свойствам более или менее широкой совокупности объектов — генеральной совокупности.

ПРИМЕР

Исходя из теории социального научения, исследователь может предположить, что демонстрация сцен насилия по телевидению ведет к повышению агрессивности подростков. Или у менеджера по кадрам некоторого крупного предприятия, исходя из собственного мнения о роли внешности, может родиться предположение, что отношение сотрудников к своим обязанностям зависит от внешнего вида руководителя. Эти совершенно различные предположения объединяет то, что их можно проверить с использованием научного подхода.

Предположение, которое проверяется с применением научного метода, будем называть **научной гипотезой**. Следует отметить, что не всякая гипотеза, а только та, которая допускает для своей проверки применение научного метода, может претендовать на научность. Кроме того, можно научно проверять гипотезы относительно любых мелких проблем, обладающих ничтожной научной или практической значимостью. *Сам факт применения научного метода вовсе не гарантирует, что проверяемая гипотеза представляет научный интерес.*

Применение научного метода для проверки гипотезы предполагает определенную последовательность действий исследователя. Исследование начинается с *операционализации* абстрактных категорий — определения операций,

[1] Представлений теоретических — в широком смысле слова, от получивших признание научных теорий до субъективных мнений. В последнем случае если гипотетические следствия окажутся состоятельными, то, кто знает, может быть родится новая научная теория.

при помощи которых соответствующие этим категориям явления (агрессивность, внешний вид и т. д.) могут быть измерены. Затем исследователь организует выборку и проводит соответствующие измерения. Результаты измерения преобразуют с использованием описательных статистик к виду, допускающему статистическую проверку научной гипотезы.

ПРИМЕР 7.1

Изучая влияние на агрессивность подростков демонстрации сцен насилия, исследователь измерил агрессивность на двух выборках, различающихся частотой и длительностью просмотра телепередач со сценами насилия. Далее он вычислил средние значения агрессивности для этих выборок: $M_1 = 6,3$ — для тех, кто чаще смотрит такие телепередачи, и $M_2 = 5,9$ — для другой выборки. Какой вывод в отношении гипотезы можно сделать на основе такого различия? Будут ли подобные различия наблюдаться в генеральной совокупности?

Любое исследование сводится к выявлению связи между переменными. В приведенном примере, в частности, исследуется связь между просмотром телепередач со сценами насилия и агрессивностью подростков. Связь эта может выражаться в величине и направлении различий между сравниваемыми группами или в знаке и величине коэффициента корреляции. То есть связь характеризуется своей силой и направлением. Однако есть еще одна не менее важная характеристика связи — ее *надежность*, «истинность».

Надежность связи непосредственно связана с репрезентативностью выборки, с тем, насколько уверенно статистики выборки позволяют судить о соответствующих параметрах генеральной совокупности. Ведь связь, обнаруженная в выборке, интересует исследователя лишь в той мере, в какой она позволяет судить о связи, которая существует в генеральной совокупности.

ПРИМЕР

Возвращаясь к предыдущему примеру 7.1, обратим внимание, что исследователь действительно обнаружил различие между выборками, свидетельствующее о справедливости его гипотезы. Однако смысл статистической проверки не в том, «различаются ли результаты двух выборок...», а в том, насколько вероятно, что существует различие между *всей* совокупностью одних и других подростков, которые могли попасть в выборку.

Надежность связи определяется тем, насколько вероятно, что обнаруженная в выборке связь будет вновь обнаружена (подтвердится) на другой аналогичной выборке, извлеченной из той же генеральной совокупности.

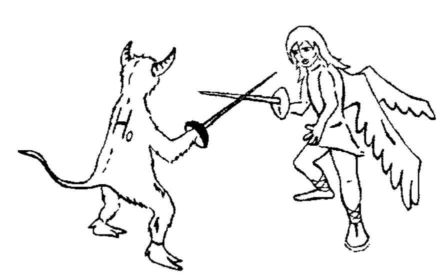
Очевидный способ проверки надежности обнаруженной в исследовании связи — это многократное проведение аналогичного исследования на разных выборках. Однако это и трудоемко и не всегда возможно. Но можно сформулировать вопрос по-другому. Если в генеральной совокупности связи нет, то какова вероятность случайного получения данного результата исследования? Иначе говоря, какова вероятность, что полученный результат является слу-

чайным, а на самом деле связи в генеральной совокупности нет? Вопрос, сформулированный таким образом, позволяет получить ответ с использованием методов статистики. Соответствующее проверяемое утверждение — об отсутствии связи — называется *статистической гипотезой.*

Статистическая гипотеза — это утверждение относительно неизвестного параметра генеральной совокупности, которое формулируется для проверки надежности связи и которое можно проверить по известным выборочным статистикам — результатам исследования. Обычно выделяют основную (нулевую) и альтернативную статистические гипотезы. **Основная (нулевая) гипотеза (H_0)** — содержит утверждение об отсутствии связи в генеральной совокупности и доступна проверке методами статистического вывода. **Альтернативная гипотеза (H_1)** — принимается при отклонении H_0 и содержит утверждение о наличии связи. При этом нулевая и альтернативная гипотезы представляют собой, в терминах теории вероятности, «полную группу несовместных событий»: если верна одна из них, то другая является ложной, и наоборот, отклонение одной из них неизбежно влечет принятие другой.

В примере 7.1 для определения надежности связи агрессивности с просмотром телепередач со сценами насилия необходимо проверить основную статистическую гипотезу H_0: $\overline{M}_1 = \overline{M}_2$ — о равенстве двух средних в генеральной совокупности (или, что то же самое, о том, что две выборки принадлежат одной генеральной совокупности). Если по результатам проверки эту гипотезу можно отклонить, то принимается альтернативная гипотеза: H_1: $\overline{M}_1 \neq \overline{M}_2$. Отклонение нулевой и принятие альтернативной статистической гипотезы в данном случае означало бы, что надежность связи достаточно велика, чтобы говорить о наличии этой связи в генеральной совокупности. Иначе говоря, это свидетельствовало бы в пользу проверяемой научной гипотезы о связи агрессивности с просмотром телепередач со сценами насилия.

Отметим, что статистическая проверка научной гипотезы следует Аристотелевой логике доказательства «от противного». Исследователь обычно заинтересован в установлении связи между изучаемыми явлениями, соответственно, его научная гипотеза обычно содержит утверждение о наличии такой связи. Но средствами статистики по результатам выборочного исследования проверяется гипотеза об отсутствии различий. И научная гипотеза подтверждается в той мере, в какой по результатам выборочного исследования возможно отклонение основной статистической гипотезы.



ПРИМЕР

Первым примером применения такой логики для проверки статистической гипотезы, по-видимому, является работа врача королевы Анны, а ранее учителя математики, Дж. Арбутнота «Довод в пользу божественного провидения, выведенный из постоянной регулярности, наблюдаемой в рождении обоих полов» (1710–

1712 гг.)[1]. В распоряжении Арбутнота были записи о рождении детей на протяжении 82 лет, которые свидетельствовали о том, что за этот период времени каждый год мальчиков рождалось больше, чем девочек. Если исходить из равновероятного рождения мальчиков и девочек ($H_0: P = 1/2$), то вероятность того, что каждый год на протяжении 82 лет мальчиков родится больше, чем девочек, составляет $(1/2)^{82} \approx 2 \cdot 10^{-25}$. Так как эта вероятность очень мала, статистическую гипотезу о равновероятном рождении мальчиков и девочек можно отклонить, приняв альтернативную гипотезу о том, что в действительности вероятность рождения мальчиков достоверно выше $1/2$. Логика обоснования «довода в пользу божественного провидения», предложенная Арбутнотом, в общих чертах сохранилась и по сей день.

ИДЕЯ ПРОВЕРКИ СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ

Рассмотрим идею проверки статистической гипотезы на примере. Предположим, психолог решил проверить пригодность разработанных ранее норм для имеющегося в его распоряжении теста интеллекта. Прежний нормативный показатель $A = 10$. На новой выборке численностью $N = 100$ человек он получил следующие результаты: $M = 10{,}6$; $\sigma = 3$.

Различия действительно обнаружены. Но интуитивно понятно, что такой результат может быть получен случайно, даже если в действительности (в генеральной совокупности) различий нет, как и наоборот, когда различия на самом деле существуют. Поэтому точный ответ в отношении генеральной совокупности по результатам выборочного исследования получить невозможно. Но методы статистики, как уже отмечалось, позволяют оценить вероятность *случайного* получения такого различия при условии, что различий на самом деле в генеральной совокупности нет (верна H_0).

В нашем примере H_0: $\overline{M}_1 = A$, то есть проверяется гипотеза, что среднее генеральной совокупности $\overline{M}$, из которой извлечена выборка, равно $A = 10$. Предположим, что выборка одного и того же объема N извлекается из такой совокупности многократно. И каждый раз вычисляется выборочное среднее значение M_x. После многократного проведения таких опытов можно построить распределение выборочных средних значений. Понятно, что выборочные средние чаще будут близки к $A = 10$, но иногда более или менее существенно отличаться от 10. Оказывается, что форма выборочного распределения для данного случая, как и для многих других, известна заранее (поэтому они называются *теоретическими распределениями*). Одна из основных теорем статистики — *центральная предельная теорема* — гласит, что распределение средних значений выборок, извлекаемых из одной и той же совокупности при достаточно большом N соответствует нормальному распределению. Среднее значение всех выборочных средних будет равно среднему значению совокуп-

---

[1] *Кендалл М., Стьюарт А.* Статистические выводы и связи. С. 687; *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. С. 247.

ности (в данном случае — $A = 10$), а дисперсия выборочных средних составит величину $m^2 = \sigma_x^2/N$, где σ_x^2 — дисперсия совокупности, N — объем каждой выборки (m еще называют *ошибкой среднего*).

Таким образом, заранее известно распределение средних для случая, когда верна H_0. Это распределение позволяет определить, насколько вероятно то или иное случайное отклонение выборочного среднего от A — среднего в генеральной совокупности. Например, из свойств нормального распределения мы знаем, что примерно 68% площади под кривой нормального распределения находится в диапазоне $\pm\sigma$ от среднего значения. Следовательно, 68% всех выборочных средних будет находиться в диапазоне $A \pm m$. Вероятность того, что выборочное среднее случайно попадет в этот диапазон составляет 0,68, а вероятность того, что оно будет отличаться от A больше чем на $1m$ составляет $1 - 0{,}68 = 0{,}32$. Аналогичным образом мы можем определить, насколько вероятно получение данного конкретного (или большего) отклонения выборочного среднего от A при условии истинности H_0.

Для нашего примера необходимо сначала определить, насколько выборочное среднее отличается от A в единицах стандартного отклонения, то есть определить соответствующее z-значение:

$$z_э = \frac{M_x - A}{\frac{\sigma_x}{\sqrt{N}}}. \qquad (7.1)$$

Формулы, подобные 7.1, позволяют получить так называемое *эмпирическое значение критерия* для соответствующего теоретического распределения (в данном случае формула 7.1 позволяет вычислить эмпирическое значение z-критерия — для нормального распределения). Подставляя выборочные значения, получаем $z = 2$. По таблице параметров нормального распределения можно определить, что в диапазоне ± 2 находится 0,954 всей площади под кривой. В соответствии с интерпретацией единичной нормальной кривой, этой площади соответствует вероятность того, что случайное отклонение от 0 будет меньше $z = \pm 2$. А для нашего случая найденная площадь соответствует вероятности того, что случайное отклонение выборочного среднего значения будет меньше $\pm(M_x - A) = \pm 0{,}6$. Соответственно, вероятность случайного отклонения выборочного среднего от генерального среднего на 0,6 и больше определяется площадью в «хвостах» под кривой нормального распределения — за пределами найденного диапазона (рис. 7.1). Следовательно, вероятность того, что данная выборка принадлежит генеральной совокупности со средним A (то есть, что верна H_0), составляет $p = 1 - 0{,}954 = 0{,}046$. Это и есть *вероятность того, что данный выборочный результат мог быть получен случайно, когда на самом деле в генеральной совокупности верна H_0* или то, что называется *p-уровнем значимости.*

Следует отметить, что выборочное распределение средних значений соответствует нормальному виду, если $N \geqslant 100$. Для выборок меньшего объема распределение средних начинает зависеть от объема выборок (точнее — от числа степеней свободы, df) и соответствует другому теоретическому распределе-

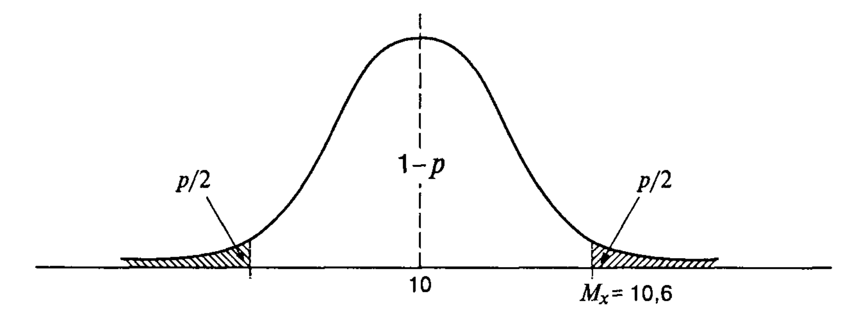


Рис. 7.1. Выборочное распределение средних значений для верной H_0

нию — *t*-Стьюдента. Тем не менее, общая последовательность проверки статистической гипотезы остается той же, как, впрочем, и для любого другого случая. Сначала вычисляется соответствующее эмпирическое значение:

$$t_э = \frac{|M - A|}{\sigma/\sqrt{N}},\ df = N - 1. \tag{7.2}$$

Затем вычисленное эмпирическое значение сопоставляется с теоретическим *t*-распределением для соответствующего числа степеней свободы *df*. Это позволяет определить *p*-уровень — вероятность того, что выборка принадлежит генеральной совокупности, для которой верна нулевая гипотеза H_0: $\overline{M} = A$.

Таким образом, в основе статистической проверки гипотез лежит представление о теоретическом распределении выборочной статистики — для условия, когда в генеральной совокупности верна нулевая статистическая гипотеза. В исследовании Арбутнота в качестве теоретического выступало биномиальное распределение для H_0: $P = {}^1/_2$, а в нашем примере — распределение выборочных средних для известной нулевой гипотезы (*Z*-распределение для больших *N* и *t*-распределение для малых *N*). В процессе проверки статистической гипотезы определяется *p*-уровень значимости (вероятность того, что нулевая статистическая гипотеза верна) путем соотнесения эмпирических значений выборочных статистик (например, разности средних) с теоретическим распределением, соответствующим нулевой статистической гипотезе.

УРОВЕНЬ СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ

Статистическая значимость (*Significant level*, сокращенно *Sig.*), или ***p*-уровень значимости** (*p-level*), — основной результат проверки статистической гипотезы. Говоря техническим языком, это вероятность получения данного результата выборочного исследования при условии, что на самом деле для генеральной совокупности верна нулевая статистическая гипотеза — то есть связи нет. Иначе говоря, это вероятность того, что обнаруженная связь носит случайный характер, а не является свойством совокупности. Именно статис-

тическая значимость, p-уровень значимости является количественной оценкой надежности связи: чем меньше эта вероятность, тем надежнее связь.

Предположим, при сравнении двух выборочных средних было получено значение уровня статистической значимости $p = 0,05$. Это значит, что проверка статистической гипотезы о равенстве средних в генеральной совокупности показала, что если она верна, то вероятность случайного появления обнаруженных различий составляет не более 5%. Иначе говоря, если бы две выборки многократно извлекались из одной и той же генеральной совокупности, то в 1 из 20 случаев обнаруживалось бы такое же или большее различие между средними этих выборок. То есть существует 5%-ная вероятность того, что обнаруженные различия носят случайный характер, а не являются свойством совокупности.

В отношении научной гипотезы уровень статистической значимости — это количественный показатель степени недоверия к выводу о наличии связи, вычисленный по результатам выборочной, эмпирической проверки этой гипотезы. *Чем меньше значение p-уровня, тем выше статистическая значимость результата исследования, подтверждающего научную гипотезу.*

Полезно знать, что влияет на уровень значимости. Уровень значимости при прочих равных условиях выше (значение p-уровня меньше), если:

- величина связи (различия) больше;
- изменчивость признака (признаков) меньше;
- объем выборки (выборок) больше.

Это демонстрируют формулы 7.1 и 7.2, как и другие формулы, предназначенные для соотнесения эмпирических значений статистик с теоретическими распределениями. В данном случае статистическая значимость возрастает (p-уровень уменьшается), когда увеличивается z-значение: при увеличении разности средних значений, при уменьшении дисперсии признака, при увеличении объема выборки.

Чем больше гипотез проверяется, тем выше шанс получить результат чисто случайно — *p-уровень увеличивается пропорционально количеству проверяемых гипотез!*

Например, если результат считается значимым при $p < 0,05$ и проверяется 20 гипотез (о корреляции или различиях), то одна из гипотез подтвердится наверняка, независимо от действительного положения дел. Единственный шанс внести ясность — проверить эти гипотезы на параллельной (идентичной) выборке.

СТАТИСТИЧЕСКИЙ КРИТЕРИЙ И ЧИСЛО СТЕПЕНЕЙ СВОБОДЫ

Статистический критерий (*Statistical Test*) — это инструмент определения уровня статистической значимости. В частности, при демонстрации логики проверки статистической гипотезы мы воспользовались z-критерием, а также

упомянули критерий *t*-Стьюдента. Как следует из логики проверки статистических гипотез, в качестве основы для применения статистических критериев используют *теоретические распределения*, для условия, когда верна нулевая гипотеза. Критерий также подразумевает *формулу*, позволяющую соотнести эмпирическое значение выборочной статистики с этим теоретическим распределением (например, формулы 7.1 и 7.2). Применяя эту формулу, исследователь вычисляет *эмпирическое значение* критерия. Полученное эмпирическое значение позволяет определить *р-уровень* — значение вероятности того, что нулевая статистическая гипотеза верна.

Помимо формулы эмпирического значения, критерий задает формулу для определения числа степеней свободы. **Число степеней свободы** (*degrees of freedom* — обозначается как *df*)— это количество возможных направлений изменчивости признака. Как правило, число степеней свободы линейно зависит от объема выборки, от числа признаков или их градаций — чем больше эти показатели, тем больше число степеней свободы. В связи с тем, что для каждого случая определение *df* имеет свою специфику, сейчас подчеркнем лишь следующее. *Каждая формула для расчета эмпирического значения критерия обязательно сопровождается правилом (формулой) для определения числа степеней свободы.*

Назначение критерия — проверка статистической гипотезы путем определения *p*-уровня значимости (вероятности того, что H_0 верна).

Выбор критерия определяется проверяемой статистической гипотезой.

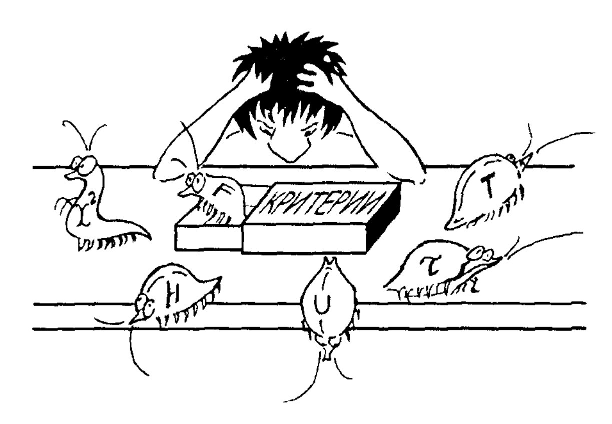
Критерий включает в себя:

- формулу расчета эмпирического значения критерия по выборочным статистикам;
- правило (формулу) определения числа степеней свободы;
- теоретическое распределение для данного числа степеней свободы;
- правило соотнесения эмпирического значения критерия с теоретическим распределением для определения вероятности того, что H_0 верна.

Для проверки статистических гипотез применяются различные критерии. При этом одному теоретическому распределению могут соответствовать разные формулы критериев — в зависимости от проверяемой статистической гипотезы. Но принцип проверки является общим для всего этого многообразия: вычисленное по формуле эмпирическое значение критерия сопоставляется с теоретическим распределением для заданного числа степеней свободы, что позволяет определить вероятность того, что H_0 верна.

ПРОВЕРКА ГИПОТЕЗ С ПОМОЩЬЮ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ

Множество разработанных статистических критериев (или статистических тестов) соответствует множеству возможных формулировок статистических гипотез. Выбор критерия представляет собой отдельную проблему, которая



Выбор критерия представляет собой отдельную проблему

будет рассматриваться нами в следующей главе. А сейчас будем исходить из того, что исследователь уже решил проблему выбора критерия, и рассмотрим общую последовательность проверки гипотезы.

При обработке данных на компьютере при помощи статистической программы (например, SPSS) исследователю достаточно указать программе, какой критерий (метод, тест) необходимо применить к заданной выборке исходных данных. Далее программа сама вычисляет эмпирическое значение критерия и сопоставляет его с теоретическим распределением. В качестве результата исследователь получает значение *р*-уровня значимости, наряду с эмпирическим значением критерия и числом степеней свободы.

Когда расчеты производятся «вручную», исследователь совершает более сложную последовательность действий для проверки гипотезы, включающую применение специальных таблиц критических значений критерия:

1. Выбор критерия в зависимости от вида исходных данных и статистической гипотезы: теоретического распределения, формул расчета эмпирического значения критерия и числа степеней свободы.
2. Расчет по исходным данным (или по имеющимся статистикам) эмпирического значения критерия и числа степеней свободы.
3. Применение «Таблицы критических значений критерия» позволяет определить значение *р*-уровня для данного числа степеней свободы.

Таблица критических значений содержит значения (квантили) теоретического распределения, соответствующие наиболее важным — критическим значениям *р*-уровня (0,1; 0,05; 0,01 и т. д.) для различных чисел степеней свободы. *р*-уровень значимости по вычисленному эмпирическому значению критерия при помощи таких таблиц определяется следующим образом. Для данного числа степеней свободы по таблице определяются ближайшие критические значения и *р*-уровни, им соответствующие. Далее значение *р*-уровня определяется в виде неравенства по правилу, которое демонстрируется на рис. 7.2 (значимость возрастает слева направо, в соответствии с убыванием *р*-уровня):

- ❒ если эмпирическое значение критерия ($K_э$) находится между двумя критическими значениями, то *р*-уровень *меньше* того критического *р*, которое находится левее;

- если $K_э$ находится левее крайнего левого критического значения (обычно это соответствует критическому $p = 0{,}1$, реже — $p = 0{,}05$), то *p*-уровень *больше*, чем крайнее правое критическое значение р;
- если $K_э$ находится правее крайнего правого критического значения, то *p*-уровень *меньше* крайнего правого критического *p*.

Например, если эмпирическое значение критерия ($K_э$) находится между $K_{0,05}$ и $K_{0,01}$, то $p < 0{,}05$. Если $K_э$ находится левее $K_{0,1}$, то $p > 0{,}1$. Если $K_э$ находится правее $K_{0,001}$, то $p < 0{,}001$.

Решение исследователя:

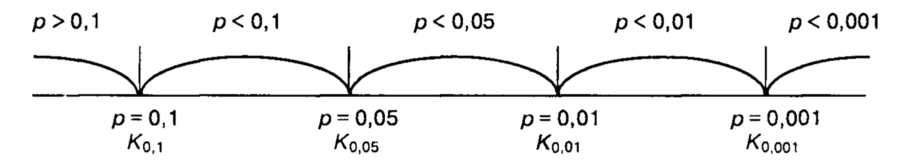


Рис. 7.2. Схема определения *p*-уровня (*p* = ... — критические значения *p*-уровня, $K_{...}$ — соответствующие критические значения критерия)

Для разных критериев возможны разные соотношения между *p*-уровнем и величиной критических его значений. Для большинства критериев (t, F, χ^2 и др.) — чем больше значение критерия, тем выше статистическая значимость (меньше *p*-уровень). Но для некоторых критериев зависимость обратная. Например, *U*-Манна-Уитни или *T*-Вилкоксона убывают по мере увеличения уровня значимости (уменьшения *p*-уровня). Тем не менее, правило остается общим, в соответствии со схемой на рис. 7.2. Например, если $t_э$ находится между $t_{0,1}$ и $t_{0,05}$ (т. е. $t_{0,1} < t_э < t_{0,05}$), то $p < 0{,}1$. И если $U_э$ находится между $U_{0,1}$ и $U_{0,05}$ (т. е. $U_{0,05} < U_э < U_{0,1}$), то $p < 0{,}1$. Если же эмпирическое значение попадает левее критического для $p = 0{,}1$ ($t_э < t_{0,1}$, но $U_э > U_{0,1}$), то уровень значимости определяется как $p > 0{,}1$.

ПРИМЕРЫ

1. Гипотеза H_0: $\overline{M} = 100$ проверяется при помощи критерия *t*-Стьюдента. Для вычисления эмпирического значения критерия $t_э$ применяется формула 7.2. На выборке $N = 36$ получены следующие значения статистик: $M = 107{,}5$, $\sigma = 15$. По формуле 7.2 $t_э = 3$, $df = 35$. Далее воспользуемся таблицей критических значений *t*-Стьюдента (приложение 2). В этой таблице строки соответствуют *df* — числам степеней свободы, столбцы — критическим значениям *p*-уровня. В строке для $df = 35$ обнаруживаем, что наше эмпирическое значение попадает в интервал между значениями 2,724 (для $p = 0{,}01$) и 3,591 (для $p = 0{,}001$). Следовательно, вероятность того, что H_0 верна, $p < 0{,}01$.
2. Предположим, та же гипотеза проверяется на выборке $N = 36$, но получены следующие значения статистик: $M = 102{,}5$, $\sigma = 15$. По формуле $t_э = 1$, $df = 35$. Воспользовавшись той же таблицей критических значений, обнаруживаем, что наше эмпирическое значение меньше, чем $t_{0,1} = 1{,}69$. Следовательно, в соответствии со схемой на рис. 7.2, $p > 0{,}1$.

СТАТИСТИЧЕСКОЕ РЕШЕНИЕ И ВЕРОЯТНОСТЬ ОШИБКИ

До сих пор под проверкой статистической гипотезы мы подразумевали процедуру определения надежности связи (*p*-уровня, как показателя статистической значимости). Однако в конечном итоге проверка статистической гипотезы должна заканчиваться *принятием статистического решения* о том, какая же гипотеза верна: нулевая — об отсутствии связи или альтернативная — о ее наличии. Соответственно, от этого зависит и окончательный, содержательный вывод исследования: подтверждена или нет исходная научная гипотеза.

Вполне очевидно, что основанием для принятия исследователем решения о том, какая гипотеза верна, является *p*-уровень — вероятность того, что верна все-таки нулевая гипотеза. Чем меньше *p*-уровень, тем с большей уверенностью можно отклонить H_0 в пользу H_1, тем самым подтвердив исходную содержательную гипотезу. Не менее очевидно и то, что, принимая решение, исследователь всегда допускает вероятность его ошибочности: ведь исследование проведено на выборке, а вывод делается в отношении генеральной совокупности. При отклонении H_0 в пользу H_1 исследователь рискует, что связи на самом деле в генеральной совокупности нет. И наоборот, решение в пользу H_0 вовсе не исключает наличие связи. Рассмотрим возможные исходы принятия решения в зависимости от действительного положения дел:

Решение исследователя:	В действительности: H_0 истинна	В действительности: H_1 истинна
Отклонить H_0 (принять H_1)	Неправильное решение, ошибка I рода, вероятность = α	Правильное решение, вероятность = $1 - \beta$ (мощность или чувствительность критерия)
Принять H_0	Правильное решение, вероятность = $1 - \alpha$ (доверительная вероятность)	Неправильное решение, ошибка II рода, вероятность = β

Как следует из таблицы, *решение исследователя зависит от того, какую* **вероятность ошибки I рода α**, *он считает допустимой*: если *p*-уровень, полученный в процессе проверки гипотезы, меньше или равен α, исследователь отклоняет H_0, и это, как правило, желательный для него результат (содержательная гипотеза подтверждается!). Отметим, что в этом случае вероятность ошибки известна, она меньше или равна α, точнее, равна *p*-уровню. Если же *p*-уровень превышает а, то принимается H_0 и содержательная гипотеза не подтверждается[1]. Но при этом **вероятность ошибки II рода β** — того, что верна все же H_1 обычно остается неизвестной.

[1] В угоду критически настроенному научному сообществу, но к огорчению исследователя!



Принятие H_0: в угоду критически настроенному научному сообществу, но к огорчению исследователя

Рассмотрим соотношение ошибок I и II рода. Предположим, как и в прошлых примерах, проверяется гипотеза об отличии среднего значения от некоторой величины *А*. Нулевой гипотезе H_0: $\overline{M} = A$ соответствует известное теоретическое распределение со средним *А*. Предположим также, что в генеральной совокупности на самом деле среднее значение больше *А* и равно *В*, а исследователь, как обычно, об этом даже и не догадывается. Этому положению дел будет соответствовать свое, «альтернативное» теоретическое распределение, сходное с распределением для H_0, но со средним *В* (рис. 7.3). На рис. 7.3 видно, что с уменьшением α растет «доверительная вероятность» $1 - \alpha$, которая определяет величину отклонения выборочного среднего от *А* для принятия H_0: уменьшая α, исследователь увеличивает возможное отклонение выборочного среднего от *А*, при котором принимается H_0. Принятие H_0 при больших отклонениях выборочного среднего от *А* увеличивает вероятность ошибки II рода, β, вероятность того, что на самом деле верна альтернативная гипотеза. Таким образом, *снижение величины α увеличивает риск допустить ошибку II рода* — не обнаружить различия или связи, которые на самом деле существуют.

Вероятность $(1 - \beta)$ называется *мощностью* (чувствительностью) *критерия*. Эта величина характеризует статистический критерий с точки зрения его способности отклонять H_0, когда она не верна. Точное значение величины мощности критерия в большинстве случаев остается неизвестным. Величина

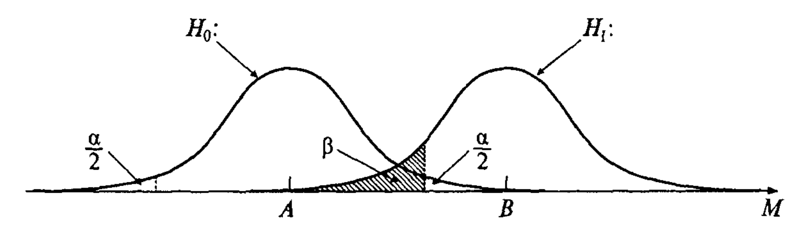


Рис. 7.3. Соотношение вероятностей ошибок I и II рода

$(1-\alpha)$ характеризует степень доверия к результатам статистической проверки и называется *доверительной вероятностью.*

Итак, основная проблема статистического вывода заключается в том, что заранее должно быть установлено оптимальное значение величины α, удовлетворяющее двум противоречивым требованиям. Величина α должна быть достаточно мала, чтобы обеспечивать доверие к результатам исследования при отклонении H_0. Величина α должна быть достаточно велика, чтобы отклонить H_0 при наличии связи (различий), не допуская ошибки II рода. Вопрос о том, какая же величина α является приемлемой, не имеет однозначного ответа. Есть лишь общие соображения, которыми можно руководствоваться при назначении α для статистического вывода:

- Для установленного значения α вероятность ошибки β уменьшается с ростом объема выборки.
- Вероятность ошибки β уменьшается при увеличении значения α (например, с 0,01 до 0,05).

Вопрос о величине α — вопрос о том, при каком же p-уровне исследователь может отклонить H_0, решается преимущественно исходя из неформальных соглашений, принятых на основе практического опыта в различных областях исследования. Традиционная интерпретация различных уровней значимости исходит из $\alpha = 0,05$ и приведена в табл. 7.1. В соответствии с ней приемлемым для отклонения H_0 признается уровень $p \leqslant 0,05$. Такая относительно высокая вероятность ошибки I рода может быть рекомендована для небольших выборок (когда высока вероятность ошибки II рода). Если объемы выборок около 100 и более объектов, то порог отклонения H_0 целесообразно снизить до $\alpha = 0,01$ и принимать решение о наличии связи (различий) при $p \leqslant 0,01$.

Таблица 7.1

Традиционная интерпретация уровней значимости при $\alpha = 0,05$

Уровень значимости	Решение	Возможный статистический вывод
$p > 0,1$	Принимается H_0	«Статистически достоверные различия не обнаружены»
$p \leqslant 0,1$	сомнения в истинности H_0, неопределенность	«Различия обнаружены на уровне статистической тенденции»
$p \leqslant 0,05$	значимость, отклонение H_0	«Обнаружены статистически достоверные (значимые) различия»
$p \leqslant 0,01$	высокая значимость, отклонение H_0	«Различия обнаружены на высоком уровне статистической значимости»











НАПРАВЛЕННЫЕ И НЕНАПРАВЛЕННЫЕ АЛЬТЕРНАТИВЫ

Основная (нулевая) статистическая гипотеза, как отмечалось, содержит утверждение о равенстве нулю (коэффициента корреляции) или о равенстве средних значений, дисперсий и т. д. Если по результатам статистической проверки основная гипотеза отклоняется, то принимается альтернативная гипотеза. Принимаемая альтернатива может быть как **направленной** (например, H_1: $r > 0$ или H_1: $M_1 > M_2$), так и **не направленной** (например, H_1: $r \neq 0$ или H_1: $M_1 \neq M_2$). То, какая альтернатива должна быть принята по результатам проверки, зависит от применяемого для проверки метода и теоретического распределения. Обычно характер альтернативы явно указывается при описании метода.

В большинстве случаев направленность или ненаправленность альтернативы зависит от формы теоретического распределения. Если оно симметрично и включает отрицательные значения, то обычно применяются ненаправленные альтернативы. Это относится к таким теоретическим распределениям, как Z-распределение (нормальное распределение), распределение t-Стьюдента и т. д. Если распределение асимметрично и может принимать только положительные значения, то применяются направленные альтернативы, например, при использовании критериев χ^2-Пирсона или F-Фишера, хотя встречаются и исключения. Важно отметить, что *выбор альтернативы — направленной или ненаправленной — исключает произвол исследователя и обычно задается выбранным методом проверки гипотезы.*

Если процедура проверки гипотезы H_0 подразумевает ненаправленную альтернативу, то критические области, соответствующие ее отклонению (принятию альтернативы), поровну распределяются по обоим «хвостам» распределения (рис. 7.4). Чаще всего интервал принятия нулевой гипотезы $(1-\alpha)$ при этом охватывает диапазон теоретических значений, симметричный относительно нуля (вспомним Z-распределение). Поэтому такие критерии часто называют *двусторонними* (*2-tailed*), имеющими «два хвоста» — для проверки ненаправленных гипотез. Заметим, что в этом случае, если принят уровень α для решения об отклонении H_0, существует *два теоретических (критических) значения*: одно отсекает $\alpha/2$ справа, а другое, отрицательное — $\alpha/2$ слева. Если проверяется направленная гипотеза, то процедура проверки допускает при-

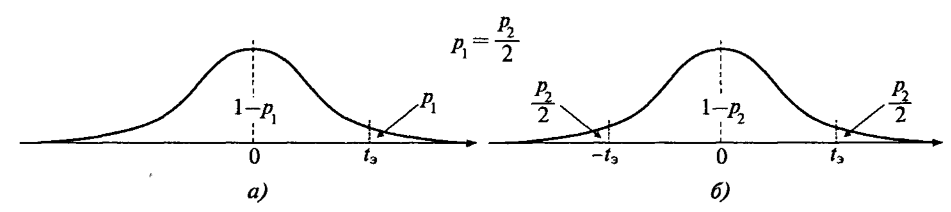


Рис. 7.4. Различие направленной *(а)* и ненаправленной *(б)* альтернатив (для одного и того же эмпирического значения p-уровень в случае *(б)* в два раза больше, чем в случае *(а)*)

нятие *односторонней альтернативы (1-tailed)* (например, H_1: $r>0$). В этом случае, если принят уровень α для решения об отклонении H_0, существует одно теоретическое (критическое) значение (для H_1: $r>0$ — положительное), и оно отсекает ровно α справа (или слева — в зависимости от направления альтернативы). Очевидно, что односторонняя альтернатива более «лояльна» к отклонению H_0 для одних и тех же выборочных результатов. *При двусторонней альтернативе, по сравнению с односторонней, нулевая гипотеза отвергается при больших значениях силы связи* (корреляции, различий средних и пр.).

Важно отметить, что принятие по результатам проверки гипотезы ненаправленной альтернативы вовсе не означает ограничение выводов лишь «ненаправленными» суждениями типа: «средние различаются», «корреляция отличается от нуля». Как следует из предыдущих рассуждений, проверка ненаправленной гипотезы является более «строгой» (при прочих равных условиях). ***Принятие ненаправленной (двусторонней) альтернативы позволяет сделать вывод о направлении связи в генеральной совокупности в соответствии с выборочными данными.***

ПРИМЕР

При проверке статистической значимости коэффициентов корреляции обычно используются ненаправленные альтернативы (H_0: $r=0$ против H_1: $r \neq 0$). Однако если H_0 отклоняется, например, при $r=-0{,}34$, то вывод не ограничивается констатацией отличия от нуля, а распространяется и на знак связи: *«обнаружена статистически достоверная отрицательная корреляция»*.

Ранее отмечалось, что определение *p*-уровня значимости — чисто техническая процедура, выполняемая компьютерной программой автоматически, а при расчетах «вручную» — по таблицам теоретических распределений (критических значений). Тем не менее, полезно знать, что существует простое соотношение между *p*-уровнями для направленных и ненаправленных альтернатив. ***Для одного и того же эмпирического значения критерия p-уровень значимости для направленной альтернативы в 2 раза меньше p-уровня для ненаправленной альтернативы.***

ПРИМЕР

Предположим, сравниваются две дисперсии. При использовании таблицы критических значений для критерия *F*-Фишера (для направленных альтернатив) (приложение 3) эмпирическое значение оказалось между критическими для $p = 0{,}05$ и $p=0{,}01$. Следовательно, для *направленной* альтернативы $p < 0{,}05$. Однако при сравнении двух дисперсий проверяется двусторонняя (ненаправленная) альтернатива, поэтому действительный уровень значимости в данном случае — $p < 0{,}1$.

Различие между направленной и ненаправленной альтернативами, кажется, еще более усложняет и без того непростую логику статистической проверки гипотез. Однако в большинстве случаев выбор альтернативы не является проблемой для исследователя — он определен самим методом (критерием) статисти-

ческой проверки и исключает возможность произвола. То, какая альтернатива предполагается, указывается явным образом при описании метода проверки. При проверке гипотезы с помощью таблиц критических значений указывается, для какой альтернативы приведены критические значения. А при использовании статистической компьютерной программы в результатах указывается, для какой альтернативы приведен *p*-уровень значимости. Например, при обработке в среде программы SPSS: Sig. (2-tailed) — *p*-уровень значимости (двусторонний), Sig. (1-tailed) — *p*-уровень значимости (односторонний).

СОДЕРЖАТЕЛЬНАЯ ИНТЕРПРЕТАЦИЯ СТАТИСТИЧЕСКОГО РЕШЕНИЯ

Статистическое решение является основанием для содержательного вывода в отношении проверяемой гипотезы. H_0 гарантирует ли отклонение H_0 истинность содержательной гипотезы о наличии связи или различий? Может ли принятие H_0 служить основанием для вывода об отсутствии связи или различий?

Принятие H_0. Из обсуждения оснований принятия статистического решения следует, что, когда принимается H_0, всегда остается вероятность того, что связь или различия все же есть. И мы ничего не можем сказать о том, насколько велика или мала эта вероятность.

Принятие H_0 не означает, что различия отсутствуют или мера связи равна нулю; из этого следует только то, что *статистически значимые результаты не обнаружены.*

Когда в результате исследования принимается H_0, *никакого содержательного вывода сделать нельзя.* Поэтому выражение «Отрицательный результат исследования — тоже результат» имеет для исследователя исключительно психотерапевтическое значение: отрицательный результат исследования — это отсутствие какого бы то ни было результата!

Отклонение H_0. В этом случае остается вероятность того, что H_0 все-таки верна и эта вероятность равна *p*-уровню значимости. Следовательно, нельзя утверждать, что результаты *доказывают* справедливость содержательной гипотезы. Корректным будет более осторожный вывод о том, что *получено свидетельство в пользу содержательной гипотезы.*

Не менее рискован содержательный вывод о причинно-следственной зависимости между изучаемыми явлениями только на основании статистической значимости связи между соответствующими признаками. Конечно, статистическая связь между признаками — это необходимое, но не достаточное условие причинно-следственной связи между ними. Утверждение о том, что явление *A* есть причина явления *B*, справедливо, если одновременно выполняются три условия (Д. Кэмпбелл, 1980): а) явления *A* и *B* статистически связаны; б) *A* происходит раньше *B*; в) отсутствует альтернативная интерпрета-



Альтернативная интерпретация статистически достоверной связи между явлениями

ция появления *В* помимо *А* (другими словами — отсутствует общая причина *С* совместной изменчивости *А* и *В*). Таким образом, применение статистических методов позволяет обосновать наличие только *статистической связи* — одного из трех признаков причинно-следственной связи.

Следует отметить, что при оформлении исследовательского отчета (курсовой или дипломной работы, публикации) статистические гипотезы и статистические решения, как правило, не приводятся. Обычно при описании результатов указывают критерий, приводят необходимые описательные статистики (средние, сигмы, корреляции и т. д.), эмпирические значения критериев, степени свободы и обязательно — *р*-уровень значимости. Затем формулируют содержательный вывод в отношении проверяемой гипотезы с указанием (обычно — в виде неравенства) достигнутого или не достигнутого уровня значимости.

ПРИМЕРЫ

На трех разных выборках проверялась содержательная гипотеза о связи креативности и тревожности. При расчете на компьютере корреляций Пирсона были получены следующие результаты для каждой из трех выборок:

1. $r = 0{,}270$; $N = 36$; $p = 0{,}11$.
2. $r = 0{,}411$; $N = 28$; $p = 0{,}02$.
3. $r = 0{,}270$; $N = 41$; $p = 0{,}08$.

Приведем примеры содержательных выводов для каждого случая:

1. Связь между креативностью и тревожностью не обнаружена ($p > 0{,}1$). Или: статистически значимой связи между креативностью и тревожностью не обнаружено ($p > 0{,}1$).
2. Обнаружена статистически значимая связь между креативностью и тревожностью ($p < 0{,}05$). Или: обнаружена статистически достоверная связь между креативностью и тревожностью ($p < 0{,}05$).
3. Связь между креативностью и тревожностью обнаружена лишь на уровне статистической тенденции ($p < 0{,}1$).

В заключении главы отметим место статистического вывода в общей последовательности проверки содержательной гипотезы.

1. Формулировка содержательной гипотезы.
2. Планирование исследования (выборка, процедура, инструментарий...), в том числе предварительная формулировка доступной проверке статистической гипотезы.

3. Проведение измерений и накопление исходных данных.
4. Окончательная формулировка статистической гипотезы, выбор статистического критерия, установление величины α — допустимой вероятности ошибки I рода.
5. Определение p-уровня статистической значимости в результате применения статистического критерия.
6. Статистический вывод: статистическое решение о принятии или отклонении H_0.
7. Формулировка содержательного вывода.

Глава 8

ВЫБОР МЕТОДА СТАТИСТИЧЕСКОГО ВЫВОДА

Приступая к операционализации содержательной гипотезы — к определению того, как будут измерены изучаемые явления, исследователь уже должен представлять себе, какому методу статистического вывода будут соответствовать получаемые в процессе исследования исходные данные. В противном случае он рискует оказаться в драматической ситуации, когда данные уже собраны, но невозможно определить метод их анализа.

Как уже отмечалось, любая содержательная гипотеза научного исследования касается связи между явлениями (свойствами, событиями) — независимо от того, содержит ли формулировка гипотезы указание на связь или на различия (между группами, условиями, событиями). Например, формулировка «мужчины и женщины различаются по коммуникативной компетентности» тождественна формулировке «коммуникативная компетентность связана с полом». Кроме того, независимо от своей формулировки, одна и та же содержательная гипотеза может быть проверена при помощи самых разных статистических методов. Ограничение на выбор статистического метода возникает только после определения того, как измерены (или будут измерены) явления, в отношении связи которых проверяется гипотеза.

ПРИМЕР

Рассмотрим некоторые возможные способы проверки одной и той же содержательной гипотезы. В одном из исследований изучалось проявление «территориального рефлекса» водителей, выезжающих с общественных автостоянок. Проверялась гипотеза о том, что водители, на место которых претендуют другие водители, покидают свое место с намеренной задержкой.[1] Гипотеза содержит утверждение о связи двух явлений: 1) интенсивность претензии на занимаемую территорию;



[1] *Солосо Р., МакЛин К.* Экспериментальная психология. СПб., 2003. С. 142.

2) интенсивность проявления «территориального рефлекса» — сопротивления претенденту. Эти явления могут быть операционализированы по-разному.

1. Наблюдение того, проявляется или нет «территориальный рефлекс» при территориальном посягательстве высокой и низкой интенсивности (два номинативных признака: а) посягательство разной интенсивности: высокой, низкой; б) проявление территориального рефлекса: есть, нет).
2. Измеряется время задержки выезда с автостоянки водителей — в зависимости от интенсивности претензии других водителей на их место (два количественных признака: интенсивность претензии, время задержки выезда).
3. Измеряется время задержки выезда с автостоянки в зависимости от характера претензии другого водителя (например, подает или нет звуковой сигнал). Одна переменная номинативная (характер претензии), другая — количественная (время задержки выезда).

Конечно, для каждого из этих случаев следует применять свой метод статистического вывода — в зависимости от измерительных шкал, в которых представлены признаки. В первом случае будут сравниваться два распределения частот проявления «территориального рефлекса»: для высокой и низкой интенсивности посягательства. Во втором случае может быть вычислена корреляция интенсивности претензии и времени задержки выезда. В третьем — речь может идти о сравнении средних значений времени выезда с автостоянки для разных случаев проявления претензии на занимаемую территорию.

Помимо типов шкал, в которых измерены или представлены изучаемые признаки, на выбор метода статистической проверки гипотезы влияет количество сравниваемых групп (градаций номинативной переменной), зависимость или независимость сравниваемых выборок и ряд других причин. Казалось бы, разнообразие способов статистической проверки должно быть очень велико и сопоставимо с бесчисленным множеством возможных содержательных гипотез. К этому можно добавить большое число разнообразных статистических критериев и вариантов их применения, которые разработаны для самых разных исследовательских ситуаций. Неудивительно, что проблема выбора метода статистического вывода, или проблема выбора критерия, зачастую становится затруднительной даже для искушенного исследователя.

Тем не менее, все бесчисленное множество содержательных гипотез может быть сведено к относительно небольшому числу типичных исследовательских ситуаций. Каждой такой ситуации соответствует своя структура исходных данных и оптимальный метод статистической проверки.

КЛАССИФИКАЦИЯ МЕТОДОВ СТАТИСТИЧЕСКОГО ВЫВОДА

Первое основание для классификации исследовательских ситуаций — это *типы шкал*, в которых измерены признаки, связь между которыми изучается. Признаки могут быть измерены либо в количественной шкале (порядковой,

Связь *X* и *Y*

Типы шкал:	I. *X*, *Y* — количественные	II. *X*, *Y* — качественные (номинативные)	III. *X* — качественный, *Y* — количественный
Задачи:	Корреляционный анализ	Анализ номинативных данных: классификаций, таблиц сопряженности, последовательностей (серий)	Сравнения выборок по уровню выраженности признака
Методы:	а) *r*-Пирсона — для метрических *X* и *Y*; б) частная корреляция и сравнение корреляций; в) *r*-Спирмена, τ-Кендалла — для ранговых *X* и *Y*	Критерий χ^2-Пирсона (для классификаций и таблиц сопряженности), критерий Мак-Нимара (для таблиц 2×2 с повторными измерениями), критерий серий (для последовательностей)	(методы сравнения) ⇩

Рис. 8.1. Классификация методов статистического вывода *о связи* двух явлений в зависимости от типа шкал, в которых они измерены

метрической), либо в качественной (номинативной) шкале. В зависимости от этого выделяются 3 ситуации (рис. 8.1).

Наиболее многочисленная группа методов относится к случаю, когда одна из переменных является количественной, а другая — качественной. Это широкий класс исследовательских ситуаций, когда задача сводится к сравнению групп (градаций номинативной переменной) по уровню выраженности признака (количественной переменной). Для решения такой задачи применяются *методы сравнения*, которые можно классифицировать по трем основаниям: а) *количество сравниваемых групп* (градаций номинативной переменной) — две или более двух; б) *соотношение сравниваемых групп*: зависимые выборки или независимые выборки; в) *шкала, в которой измерен количественный признак*: метрическая, ранговая. Таким образом, можно выделить 8 основных методов сравнения (рис. 8.2).

Методы сравнения (*X* — качественный, *Y* — количественный)

Количество выборок (градаций *X*)		Две выборки		Больше двух выборок	
Зависимость выборок		Независимые	Зависимые	Независимые	Зависимые
Признак *Y*	метрический	Параметрические методы сравнения			
		t-Стьюдента, для независимых выборок	*t*-Стьюдента, для зависимых выборок	ANOVA	ANOVA, с повторными измерениями
	ранговый	Непараметрические методы сравнения			
		U-Манна-Уитни, критерий серий	*T*-Вилкоксона, критерий знаков	*H*-Краскала-Уоллеса	χ^2-Фридмана

Рис. 8.2. Классификация методов статистического вывода *о различии* выборок по уровню выраженности количественного признака

МЕТОДЫ КОРРЕЛЯЦИОННОГО АНАЛИЗА

Проверяемая H_0: коэффициент корреляции равен нулю.

Условие применения: а) два признака измерены в ранговой или метрической шкале на одной и той же выборке; б) связь между признаками является монотонной (не меняет направления по мере увеличения значений одного из признаков).

Структура исходных данных:

№	*X*	*Y*
1	1	3
2	3	5
3	3	1
4	8	2
...	...	...
N	9	16

Обычно изучаются корреляции между множеством *P* переменных. В таком случае вычисляются корреляции между всеми возможными парами этих переменных. Результатом является *корреляционная матрица*, включающая $P(P-1)/2$ значений коэффициентов парной корреляции. Под корреляционным анализом обычно и понимают изучение связей по корреляционной матрице.

Методы:

Корреляция *r*-Пирсона — для метрических переменных.

Условие применения: а) распределения *X* и *Y* существенно не отличаются от нормального.

Дополнительно: частная корреляция — для изучения зависимости корреляции *X* и *Y* от влияния переменной *Z*; сравнение корреляций — для независимых и зависимых выборок.

Корреляции *r*-Спирмена, τ-Кендалла — для порядковых переменных.

МЕТОДЫ АНАЛИЗА НОМИНАТИВНЫХ ДАННЫХ

В зависимости от цели исследования и структуры исходных данных выделяются три группы методов, соответствующих решаемым задачам:

- анализ классификаций;
- анализ таблиц сопряженности;
- анализ последовательностей (серий).

Анализ классификаций

Условие применения: для каждого объекта (испытуемого) выборки определена его принадлежность к одной из категорий (градаций) *X* (получено эмпирическое распределение объектов по *X*); известно теоретическое (ожидаемое) распределение по *X* (обычно — равномерное).

ПРИМЕР

Исследовались различия в предпочтении респондентами пяти политических лидеров. H_0: эмпирическое распределение предпочтений респондентов не отличается от равномерного.

Таблица сопряженности:

Полит. лидер (*X*)	Распределение	
	Эмпирическое	Теоретическое
1	16	21
2	37	21
3	29	21
4	13	21
5	10	21
Всего:	105	105

Проверяемая H_0: эмпирическое (наблюдаемое) распределение *X* не отличается от теоретического (ожидаемого).

Метод: критерий χ^2-Пирсона.

Анализ таблиц сопряженности

Условие применения: для каждого объекта (испытуемого) выборки определена его принадлежность к одной из категорий (градаций) *X* и к одной из категорий (градаций) *Y* (получена перекрестная классификация объектов по двум основаниям — *X* и *Y*).

Следует различать *три ситуации* — в зависимости от числа градаций и соотношения *X* и *Y*:

- ❒ число градаций *X* и (или) *Y* больше двух (общий случай);
- ❒ таблицы сопряженности 2×2 с независимыми выборками;
- ❒ таблицы сопряженности 2×2 с повторными измерениями.

Общий случай: число градаций больше двух

ПРИМЕР

Исследовались различия между мужчинами и женщинами в предпочтениях пяти политических лидеров.

Структура исходных данных:

№	X (пол)	Y (политический лидер)
1	1	3
2	2	5
...	...	...
N	1	2

Таблица сопряженности:

		Y (политический лидер)				
		1	2	3	4	5
X (пол)	муж. (1)	7	22	11	5	6
	жен. (2)	9	15	18	8	4
Всего:		16	37	29	13	10

Проверяемая H_0: два вида классификации (X и Y) являются независимыми.
Метод: критерий χ^2-Пирсона.

Таблицы сопряженности 2×2 с независимыми выборками

ПРИМЕР

Методом «потерянных писем» исследовалась склонность людей передавать хорошие и плохие новости. Из 60 открыток с «хорошими» новостями до адресата дошли 35, а из 120 с «плохими» новостями дошли 23 открытки. Действительно ли люди более склонны передавать хорошие новости, чем плохие?

Таблица сопряженности (2×2):

		Y (открытки)	
		не отправленные	отправленные
X (новость)	плохая	97	23
	хорошая	25	35

Проверяемая H_0: два вида классификации (X и Y) являются независимыми.
Методы: критерий χ^2-Пирсона (с поправкой на непрерывность Йетса), точный критерий Фишера.

Таблицы сопряженности 2×2 с повторными измерениями

ПРИМЕР

Необходимо сравнить два вопроса, заданных одной и той же группе испытуемых, по соотношению ответов «да» и «нет»:

		Y (вопрос 2)	
		«Да»	«Нет»
X (вопрос 1)	«Да»	$a = 40$	$b = 30$
	«Нет»	$c = 15$	$d = 20$

Проверяемые H_0: а) $a = d$; б) $c = b$.

Метод: соотнесение диагональных элементов таблицы 2×2 при помощи метода Мак-Нимара (по критерию z или χ^2).

Анализ последовательности (серий)

Условие применения: объекты упорядочены (по времени или по уровню выраженности признака); каждый объект отнесен к одной из двух категорий (*X* или *Y*).

ПРИМЕРЫ

Случай 1. События *X* и *Y* чередуются следующим образом:
X X X X X Y Y Y Y Y X X Y Y Y Y Y X Y Y

Случай 2. Значения количественного признака, измеренного для выборки *X* и для выборки *Y*, после ранжирования чередуются следующим образом:
X X X X X Y Y Y Y Y X X Y Y Y Y Y X Y Y

Проверяемые H_0: события *X* распределены среди событий *Y* случайно (случай 1); выборки *X* и *Y* не различаются по распределению значений количественного признака (случай 2).

Метод: критерий серий.

МЕТОДЫ СРАВНЕНИЯ ВЫБОРОК ПО УРОВНЮ ВЫРАЖЕННОСТИ ПРИЗНАКА

В зависимости от решаемых задач методы внутри этой группы классифицируются по трем основаниям:

- Количество градаций *X*:
  - а) сравниваются 2 выборки;
  - б) сравниваются больше 2 выборок.
- Зависимость выборок:
  - а) сравниваемые выборки независимы;
  - б) сравниваемые выборки зависимы.
- Шкала *Y*:
  - а) *Y* — ранговая переменная;
  - б) *Y* — метрическая переменная.

По последнему основанию методы делятся на две большие группы: параметрические методы (критерии) — для метрических переменных и непараметрические методы (критерии) — для порядковых (ранговых) переменных. *Параметрические методы* проверяют гипотезы относительно *параметров распределения* (средних значений и дисперсий) и основаны на предположении о

нормальном распределении в генеральной совокупности. *Непараметрические методы* не зависят от предположений о характере распределения и не касаются параметров этого распределения.

Сравнение двух выборок

Проверяемая H_0: две совокупности (которым соответствуют выборки) не отличаются по уровню выраженности измеренного признака.

Сравнение двух независимых выборок

Условия применения: признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из двух независимых выборок.

ПРИМЕР

Исследование различий между юношами и девушками по тревожности, измеренной в количественной (ранговой, метрической) шкале.

Структура данных:

№	*X* (пол)	*Y* (тревожность)
1	1	10
2	2	9
3	2	3
4	1	8
...	...	...
N	1	6

Методы:

***Y* — метрическая переменная:** сравнение двух средних значений (параметрический критерий *t*-Стьюдента для независимых выборок).

Условия применения: признак измерен в (а) метрической шкале, (б) дисперсии двух выборок гомогенны (статистически достоверно не различаются). Если не выполняется хотя бы одно из этих условий, то применяется непараметрический критерий *U*-Манна-Уитни.

Дополнительно: возможно сравнение двух дисперсий (параметрический критерий *F*-Фишера).

***Y* — ранговая (порядковая) переменная:** сравнение двух независимых выборок по уровню выраженности порядковой или бинарной переменной (критерий *U*-Манна-Уитни, критерий серий).

Сравнение 2-х зависимых выборок

Условия применения: (а) признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из двух зависимых выборок: либо при-

знак измерен дважды на одной и той же выборке, либо каждому испытуемому из одной выборки поставлен в соответствие по определенному критерию испытуемый из другой выборки; (б) измерения положительно коррелируют. Если эти условия не выполняются, то выборки следует признать независимыми.

ПРИМЕРЫ

1. Изучался эффект социально-психологического тренинга. Каждому ученику класса (численностью N) задавался вопрос: «Как часто твое мнение совпадает с мнением твоих одноклассников», отвечать на который предлагалось при помощи 10-балльной шкалы. Ученики отвечали на вопрос дважды: до (X_1) и после (X_2) тренинга.

Структура данных:

№	X_1	X_2
1	8	10
2	8	9
3	3	4
4	5	5
...	...	...
N	6	7

2. Изучалось различие в самооценке единства мнений в супружеских парах (всего N пар) между мужьями и их женами. Для этого на вопрос «Как часто Ваше мнение совпадает с мнением супруги (супруга)» при помощи 10-балльной шкалы отвечали мужья каждой пары (X_1) и их жены (X_2).

Структура данных та же, что и для предыдущего примера, но № — номер пары.

Методы:

Y — **метрическая переменная:** сравнение двух средних значений (параметрический критерий t-Стьюдента для зависимых выборок).

Условие применения: признак измерен в метрической шкале. Если это условие не выполняется, то применяется непараметрический критерий T-Вилкоксона.

Y — **ранговая (порядковая) переменная:** сравнение двух зависимых выборок по уровню выраженности порядковой или бинарной переменной (критерий T-Вилкоксона, критерий знаков).

Сравнение более двух выборок

Проверяемая H_0: несколько совокупностей (которым соответствуют выборки) не отличаются по уровню выраженности измеренного признака.

Сравнение более двух независимых выборок

Условия применения: признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из k независимых выборок ($k > 2$).

ПРИМЕР

Исследовалось влияние интервала между 5 повторениями вербального материала на продуктивность (Y) последующего его воспроизведения. Интервал между повторениями (X — три градации) составил: для 1 группы — 0 мин; для 2 группы — 30 мин, для 3 группы — 60 мин.

Структура данных:

№	X (интервал)	Y (эффективность воспроизведения)
1	1	8
2	3	9
3	2	4
4	1	5
...	...	...
N	2	6

Методы:

***Y* — метрическая переменная:** дисперсионный анализ (ANOVA) для независимых выборок (параметрический метод).

Дополнение: метод допускает сравнение выборок более чем по одному основанию — когда деление на выборки производится по нескольким номинативным переменным, каждая из которых имеет 2 и более градаций.

ПРИМЕР

Исследовалось влияние на продуктивность воспроизведения (Y) вербального материала: а) интервала между повторениями (X_1 — 3 градации) и б) объема материала (X_2 — 2 градации).

Структура данных:

№	X_1 (интервал)	X_2 (объем)	Y (эффективность воспроизведения)
1	1	2	8
2	3	2	9
3	2	1	4
4	1	1	5
...	...		...
N	2	2	6

Условия применения: признак Y измерен в (а) метрической шкале, (б) дисперсии выборок гомогенны (статистически достоверно не различаются). Если не выполняется хотя бы одно из этих условий, то:

***Y* — ранговая (порядковая) переменная:** сравнение более двух независимых выборок по уровню выраженности ранговой переменной (непараметрический критерий *Н*-Краскала-Уоллеса).

Ограничение: метод позволяет сравнивать выборки только по одному основанию, когда деление на группы производится по одной номинативной переменной, имеющей более 2-х градаций.

Сравнение более двух зависимых выборок

Условия применения: (а) признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из k зависимых выборок ($k > 2$): как правило, признак измерен несколько раз на одной и той же выборке; (б) измерения положительно коррелируют.

ПРИМЕРЫ

Исследовалось влияние положения элементов в ряду (переменная X, 3 градации: начало, середина, конец ряда) на продуктивность их воспроизведения каждым из N испытуемых (переменная Y: доля воспроизведенных элементов).

Структура данных:

№	Y_1	Y_2	Y_3
1	0,1	0,3	0,4
2	0,3	0,1	0,2
3	0,2	0,1	0,5
4	0,1	0,2	0,3
...	...	...	...
N	0,2	0,1	0,2

Методы:

***Y* — метрическая переменная:** дисперсионный анализ (ANOVA) с повторными измерениями (параметрический метод).

Дополнение: метод допускает сравнение выборок более чем по одному основанию — когда помимо деления на зависимые выборки, вводятся номинативные переменные, которые имеют 2 и более градаций и делят испытуемых на независимые выборки.

ПРИМЕР

Исследовалось влияние на продуктивность воспроизведения (переменная Y: доля воспроизведенных элементов): а) положения элементов в ряду (переменная X_1, 3 градации: начало, середина, конец ряда); б) способа предъявления ряда (переменная X_2, 2 градации).

Структура данных:

№	X_2	Y_1	Y_2	Y_3
1	1	0,1	0,3	0,4
2	2	0,3	0,1	0,2
3	1	0,2	0,1	0,5
4	1	0,1	0,2	0,3
...		...	...	...
N	2	0,2	0,1	0,2

Условия применения: а) признак *Y* измерен в метрической шкале; б) дисперсии сравниваемых выборок гомогенны (статистически достоверно не различаются). Если не выполняется хотя бы одно из этих условий, то:

***Y* — ранговая (порядковая) переменная:** сравнение более двух зависимых выборок по уровню выраженности ранговой переменной (непараметрический критерий χ^2-Фридмана).

Ограничение: метод позволяет сравнивать зависимые выборки только по одному основанию — повторным измерениям.

Глава 9

АНАЛИЗ НОМИНАТИВНЫХ ДАННЫХ

Методы, о которых пойдет речь в этой главе, касаются проверки, по-видимому, самого широкого класса гипотез — в отношении тех явлений, измерения которых доступны в номинативной шкале.

ПРИМЕРЫ

Кто чаще обращается в службу знакомств: мужчины или женщины?

Зависит ли количество аварий на производстве от дня недели?

Можно ли утверждать, что водители-женщины чаще становятся участниками ДТП (дорожно-транспортных происшествий)?

Можно ли утверждать, что выигрыши в игре распределены не случайно среди проигрышей?

Данные для ответов на подобные обыденные и чисто академические вопросы могут быть получены при помощи простого способа — классификации событий и людей по интересующим градациям. И несмотря на, казалось бы, бесчисленное многообразие подобных ситуаций, все они могут быть сведены *к трем типичным случаям*:

1 — сравнение наблюдаемого (эмпирического) распределения частот с ожидаемым (теоретическим) распределением;

2 — сравнение двух или более наблюдаемых распределений частот;

3 — сравнение наблюдаемого распределения событий *X* среди событий *Y* (серий *X*, *Y*) со случайным распределением.

ПРИМЕРЫ

Случай I.

1. Кто чаще обращается в службу знакомств: мужчины или женщины? Для ответа на этот вопрос необходимо: а) подсчитать количество женщин и мужчин, обратившихся в службу знакомств; б) воспользовавшись методом статистической проверки, сопоставить полученное эмпирическое соотношение мужчин и женщин с ожидаемым (теоретическим) равномерным распределением.
2. Зависит ли количество аварий на производстве от дня недели? Проверка этого предположения требует выполнения сходных действий: а) подсчитать количество аварий для каждого дня недели за достаточно длительный промежуток вре-

мени; б) воспользовавшись методом статистической проверки, сопоставить полученное эмпирическое распределение количества аварий по дням недели с ожидаемым (теоретическим) равномерным распределением.

Случай II.

1. Зависит ли предпочтение напитка (минеральная вода, сок, лимонад) от сезона (зима, весна, лето, осень)? Для проверки этого предположения необходимо для каждого респондента определить тип предпочитаемого напитка (первая номинативная переменная, 3 градации) и сезон опроса (вторая номинативная переменная — 4 градации).



2. Зависит ли предпочтение одного из пяти кандидатов на выборах от пола потенциального избирателя? Для проверки этого предположения необходимо для каждого респондента определить пол (первая номинативная переменная, 2 градации) и предпочитаемого кандидата, одного из пяти (вторая номинативная переменная, 5 градаций).
3. Повлияла ли рекламная кампания на выбор респондентами одного из двух товаров? Это предположение требует опроса респондентов на предмет предпочтения одного из двух товаров дважды: до рекламной кампании (первая номинативная переменная, две градации) и после нее (вторая номинативная переменная, те же две градации).

Для решения подобных задач, связанных с анализом классификаций или таблиц сопряженности, оказывается достаточным применение одного и того же критерия — χ^2-Пирсона:

$$\chi^2_э = \sum_{i=1}^{P} \frac{(f_э - f_т)^2}{f_т}, \quad df = (k-1)(l-1), \qquad (9.1)$$

где P — количество ячеек таблицы распределения или сопряженности, содержащих эмпирические значения частот; $f_э$, $f_т$ — эмпирическое и теоретическое значения частот для одной ячейки; k — число градаций сопоставляемых распределений; l — количество сопоставляемых распределений. Приведенная формула является общей для различных ситуаций, и в каждом случае ее применение обладает своей спецификой.

ПРИМЕРЫ

Случай III.

1. Является ли закономерным последовательный повтор выигрышей среди проигрышей в игре или это случайные совпадения?
2. В последовательности событий X и Y является ли закономерным их чередование (X после Y и наоборот)?
3. Наблюдается ли закономерность в чередовании быстрых и медленных реакций на некоторый стимул: имеют ли они тенденцию к группированию или после медленной реакции следует быстрая (и наоборот)?

Для решения задач такого типа необходимо упорядочить события во времени и подсчитать число серий. *Серия* — это последовательность однотип-

ных событий, непосредственно перед и после которой произошли события другого типа. Далее применяется *критерий серий*, позволяющий определить вероятность случайного появления наблюдаемого числа серий при условии хаотичного распределения событий *X* среди событий *Y*.

Очень часто при исследовании классификаций, сопряженности или последовательности нет необходимости в накоплении данных в привычных таблицах типа «объект-признак»: результаты наблюдений сразу заносят в таблицу распределения (сопряженности) или составляют последовательность. В этом случае нет необходимости в использовании специальных статистических программ, и все расчеты можно провести «вручную». Тем более что они не составляют особого труда.

АНАЛИЗ КЛАССИФИКАЦИИ: СРАВНЕНИЕ ЭМПИРИЧЕСКОГО И ТЕОРЕТИЧЕСКОГО РАСПРЕДЕЛЕНИЙ

Две градации

Эта задача сводится к сравнению численности двух долей объектов (людей, событий и т. д.) в совокупности: обладающих и не обладающих некоторым свойством.

ПРИМЕР

Мы можем сопоставлять долю мужчин, которым больше нравятся блондинки, с долей мужчин, которым больше нравятся девушки с темными волосами. Или сопоставлять доли голосующих «за» и «против» введения моратория на смертную казнь.

Обычно, сопоставляя доли, мы надеемся обнаружить различия их пропорции от некоторого ожидаемого соотношения. Соотношение численности групп, которое мы получаем в результате исследования, называется *эмпирическим распределением*. Ожидаемому соотношению соответствует *теоретическое распределение*. В качестве теоретического распределения чаще всего выступает равномерное распределение.

Изучая отношение людей к введению моратория на смертную казнь, мы надеемся, что численность группы голосующих «за» будет отличаться от численности группы голосующих «против», то есть распределение голосующих на две категории будет отличаться от равномерного распределения.

Формулировка проверяемой H_0: соотношение долей в генеральной совокупности не отличается от ожидаемого (теоретического) соотношения.

Исходные данные: определена принадлежность каждого испытуемого к одной из двух категорий номинативной переменной. Задано ожидаемое (теоретическое) соотношение численности категорий.

Эта гипотеза проверяется при помощи формулы 9.1 для критерия χ^2, где $P = 2$ (сумма состоит из двух слагаемых), $k = 2$, $l = 2$, каждая из двух эмпирических частот соответствует численности сравниваемых групп. Численности каждой из сравниваемых групп (эмпирической частоте) ставится в соответствие теоретическая частота. Сумма теоретических частот равна сумме эмпирических частот, а соотношение теоретических частот равно ожидаемому (теоретическому) соотношению.

Следует отметить, что точное решение для такого рода задач дает применение *биномиального критерия*. Но поскольку его расчет трудоемок, а таблицы критических значений громоздки, мы предлагаем для расчетов «вручную» использовать приближение при помощи критерия χ^2. При расчетах на компьютере в подобных случаях все же следует предпочесть биномиальный критерий (см. раздел «Обработка на компьютере»).

ПРИМЕР 9.1

А) Из 50 опрошенных по поводу отношения к введению моратория на смертную казнь 30 были «за», 20 — «против» (предполагается, что выборка репрезентативна генеральной совокупности). Можно ли утверждать на основании этого опроса, что в совокупности количество сторонников превышает количество противников введения моратория на смертную казнь?

	Распределение:	
	эмпирическое	теоретическое
«За»	30	25
«Против»	20	25
Сумма:	50	50

Шаг 1. Формулируем H_0: сравниваемые доли равны между собой (эмпирическое распределение соответствует равномерному распределению).

Шаг 2. Выбираем для принятия статистического решения $\alpha = 0{,}05$.

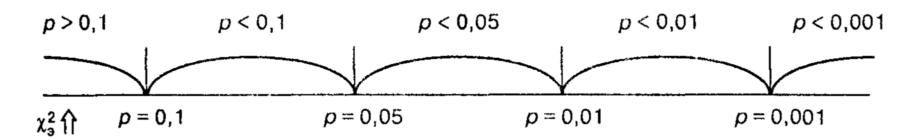
Шаг 3. Вычисляем эмпирическое значение критерия. Задача сводится к сопоставлению эмпирического распределения 30 : 20 с идентичным по общей численности, но равномерным теоретическим распределением 25 : 25. Следовательно:

$$(f_э)_1 = 30;\ (f_э)_2 = 20;\ (f_т)_1 = 25;\ (f_т)_2 = 25.$$

Подставляем эти значения в формулу 9.1:

$$\chi^2_э = \frac{(30-25)^2}{25} + \frac{(20-25)^2}{25} = 1 + 1 = 2,\ df = 1.$$

Шаг 4. Определяем p-уровень. По таблице критических значений теоретического распределения χ^2-Пирсона (приложение 4) для $df = 1$ видим, что наше эмпирическое значение $\chi^2_э$ находится левее критического значения для $p = 0{,}1$:



Шаг 5. Принимаем статистическое решение. В соответствии со схемой определения p-уровня $p > 0{,}1$, и мы не можем отклонить H_0, так как $p > \alpha$.

Шаг 6. Формулируем содержательный вывод. В результате исследования не обнаружены статистически значимые различия в соотношении численности сторонников и противников введения моратория на смертную казнь ($p > 0{,}1$). Или: численность сторонников и противников введения моратория на смертную казнь статистически значимо не различается ($p > 0{,}1$).

Б) Предположим теперь, что было опрошено не 50, а 100 человек, и соотношение высказавшихся «за» и «против» сохранилось. Тогда эмпирические частоты составили бы 60 «за» и 40 «против», а соответствующие теоретические частоты равнялись бы 50. Число степеней свободы не меняется, а эмпирическое значение критерия увеличивается: $\chi^2_э = 4$. В соответствии с таблицей критических значений χ^2 и со схемой определения p-уровня $p < 0{,}05$, и мы можем отклонить H_0, так как $p < \alpha$. Тогда содержательный вывод будет другим: численность сторонников введения моратория на смертную казнь статистически достоверно выше численности противников введения моратория ($p < 0{,}05$).

Обратите внимание: принятие H_0 не позволяет сделать никакого вывода о соотношении численности сравниваемых групп. Напротив, отклонение H_0 позволяет в данном случае говорить не только о различии сравниваемых долей, но и о направлении различий — о том, что одна доля больше другой.

Отметим, что в качестве ожидаемого (теоретического) распределения может выступать не обязательно равномерное распределение. Например, мы можем проверять содержательную гипотезу о том, что некоторая группа составляет по численности менее 20% совокупности. Тогда соотношение теоретических частот будет не 1:1, как в рассмотренном примере, а 1: 4. В остальном весь ход решения остается прежним.

ПРИМЕР 9.2

Рассмотрим исследование, в котором проводилось сравнение частоты рождения мальчиков в индейских семьях английского города, где подавляющую часть населения составляли выходцы из Америки[1]. Средняя частота рождения мальчиков в Англии составляет 52%, а в данном случае за период наблюдения из 20 родившихся детей мальчиков оказалось 5. Можно ли на этом основании сделать вывод о том, что в индейских семьях этого города мальчики рождаются достоверно реже, чем в целом по Англии?

Шаг 1. Формулируем H_0: $P = 0{,}52$ (выборочные данные согласуются с вероятностью рождения мальчиков $P = 0{,}52$).

---

[1] Справочник по прикладной статистике. В 2 т. / Под ред. Э. Ллойда, У. Ледермана, Ю. Тюрина. М., 1989. С. 212.

Шаг 2. Выбираем для принятия статистического решения $\alpha = 0{,}05$.

Шаг 3. Составляем таблицу эмпирических и теоретических частот и вычисляем эмпирическое значение критерия.

	Распределения:	
	эмпирическое	теоретическое
Мальчики	5	10,4
Девочки	15	9,6
Сумма:	20	20

Задача сводится к сопоставлению эмпирического распределения 5:15 с идентичным по общей численности теоретическим распределением (0,52:0,48). Следовательно:

$$(f_э)_1 = 5;\ (f_э)_2 = 15;\ (f_т)_1 = 10{,}4;\ (f_т)_2 = 9{,}6.$$

Подставляем эти значения в формулу 9.1:

$$\chi^2_э = \frac{(5-10{,}4)^2}{10{,}4} + \frac{(15-9{,}6)^2}{9{,}6} = 5{,}84,\ df = 1.$$

Шаг 4. Определяем p-уровень. По таблице критических значений теоретического распределения χ^2-Пирсона (приложение 4) для $df = 1$ видим, что наше эмпирическое значение $\chi^2_э$ находится между критическими значениями для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$.

Шаг 5. Принимаем статистическое решение. Так как $p < \alpha$, то H_0 можно отклонить.

Шаг 6. Формулируем содержательный вывод. В индейских семьях этого города мальчики действительно рождаются достоверно реже, чем в целом по Англии ($p < 0{,}05$).

Обработка на компьютере: биномиальный критерий

Исходные данные: значения бинарной номинативной переменной (0, 1) определены для каждого члена выборки и представлены одним столбцом.

Выбираем: **Analyze** (Метод) > **Nonparametric tests...** (Непараметрические методы) > **Binomial...** (Биномиальный). В открывшемся окне диалога переносим необходимую бинарную переменную из левого в правое окно (**Test Variable List**), переменных может быть несколько.

Если теоретическое распределение является равномерным, то нажимаем OK и получаем результаты.

Если теоретическое распределение не является равномерным, то необходимо задать ожидаемые (теоретические) пропорции (доли) для той градации, которая встречается в данных раньше. Для этого в окне **Test proportion** (Ожидаемая пропорция) вводим ожидаемую долю для градации. Нажимаем OK и получаем результаты.

Результаты (для данных примера 9.2)

Binomial Test

		Category	N	Observed Prop.	Test Prop.	Exact Sig. (1-tailed)
var	Group 1	1.00	5	.25	.52	.013(a)
	Group 2	.00	15	.75		
	Total		20	1.00		

a Alternative hypothesis states that the proportion of cases in the first group < .52.

Observed Prop. — наблюдаемая доля для каждой категории (Category); Test Prop. — теоретическая доля для первой из категорий; Exact Sig. (1-tailed) — точное значение *p*-уровня для односторонней альтернативы (направленной гипотезы).

Примечание. Если проверяется ненаправленная гипотеза, то полученное значение *p*-уровня необходимо умножить на 2.

Более двух градаций

Как и в предыдущем случае, при сопоставлении нескольких градаций чаще всего проверяют гипотезу о том, различаются ли по численности соответствующие доли совокупности. Это соответствует задаче сопоставления эмпирического и равномерного теоретического распределения. Но ожидаемое (теоретическое) распределение может быть и любым другим: последовательность решения при этом не меняется. Для проверки подобных гипотез применяют критерий χ^2-Пирсона (формула 9.1), который еще называют *критерием согласия* (эмпирического и теоретического распределений).

ПРИМЕР 9.3

С целью предсказания результатов выборов исследовалось предпочтение потенциальными избирателями пяти политических лидеров. По результатам опроса репрезентативной выборки из 120 респондентов была составлена таблица распределения их предпочтений:

Политические лидеры:	1	2	3	4	5
Кол-во «поклонников»:	21	37	29	15	18

Можно ли утверждать, что в совокупности всех потенциальных избирателей наблюдаются существенные различия в соотношении предпочтений пяти политических лидеров? Иначе говоря, отличается ли распределение предпочтений потенциальных избирателей от равномерного распределения?

Отметим, что в отношении данной группы респондентов ответ очевиден: да, предпочтения распределены явно не равномерно. Но вопрос при статистической проверке формулируется иначе: можно ли распространить этот вывод на генеральную совокупность, из которой извлечена данная выборка респондентов?

Поскольку $N > 100$, выбираем для принятия статистического решения $\alpha = 0{,}01$.

H_0: эмпирическое распределение соответствует теоретическому равномерному распределению. Задача сводится к сопоставлению эмпирического распределения с идентичным по общей численности, но равномерным теоретическим (ожидаемым) распределением:

Политические лидеры	Распределение предпочтений:	
	эмпирическое	теоретическое
1	21	24
2	37	24
3	29	24
4	15	24
5	18	24
Всего	120	120

По формуле 9.1 число слагаемых $P = 5$, $k = 5$, $l = 2$, $df = 4$.

$$\chi^2_э = \frac{(21-24)^2}{24} + \frac{(37-24)^2}{24} + \frac{(29-24)^2}{24} + \frac{(15-24)^2}{24} + \frac{(18-24)^2}{24} = 13{,}333.$$

По таблице критических значений теоретического распределения χ^2-Пирсона (Приложение 4) для $df = 4$ видим, что наше эмпирическое значение $\chi^2_э$ меньше критического значения для $p = 0{,}01$. Следовательно, в соответствии со схемой определения p-уровня для данного случая $p < 0{,}01$. Так как $p < \alpha$, то принимаем *статистическое решение*: отклоняется нулевая гипотеза о соответствии распределения предпочтений в генеральной совокупности равномерному распределению. Таким образом, корректен следующий *содержательный вывод: обнаружены различия в предпочтениях потенциальными избирателями пяти политических лидеров ($p < 0{,}01$).*

Отметим, что в этом случае, отклоняя H_0, мы принимаем альтернативную гипотезу о том, что распределение предпочтений является неравномерным. ***Но альтернативная гипотеза не содержит и не может содержать утверждения о том, что в какой-то конкретной ячейке наблюдений больше, а в какой-то меньше.*** Любая конкретизация этого утверждения будет некорректной. Для утверждений о том, что в какой-то ячейке (градации) наблюдений больше или меньше, необходима дополнительная статистическая проверка.

Например, на первый взгляд справедливое утверждение о том, что лидер № 2 предпочитается чаще, чем лидер № 3 (пример 9.3), при дополнительной статистической проверке не подтверждается. Сравнение распределения 37:29 с ожидаемым равномерным распределением 33:33 дает: $\chi^2_э = 0{,}970$; $df = 1$. Величина эмпирического значения критерия меньше критического значения для $df = 1$, $p = 0{,}1$ (эмпирическое значение располагается левее критического значения критерия для $p = 0{,}1$). Следовательно, в данном случае $p > 0{,}1$, H_0 не отклоняется: не обнаружены различия в предпочтениях двух политических лидеров ($p > 0{,}1$).

Подобная *проблема множественных сравнений* возникает всегда, если нулевая гипотеза содержит утверждение о равенстве более чем двух величин. При ее отклонении принимается альтернативная гипотеза, содержащая изрядную долю неопределенности: сравниваемые величины не тождественны. Для кон-

кретизации этого утверждения необходимы, как правило, парные сравнения величин, в отношении которых проверяется гипотеза.

Обработка на компьютере: критерий согласия χ^2

Исходные данные: значения номинативной переменной (более 2-х градаций) определены для каждого члена выборки и представлены одним столбцом.

Выбираем: **Analyze** (Метод) > **Nonparametric tests** (Непараметрические методы) > **Chi-square...** (Хи-квадрат). В открывшемся окне диалога переносим необходимую переменную из левого в правое окно (**Test Variable List**), переменных может быть несколько.

Если теоретическое распределение является равномерным, то нажимаем OK и получаем результаты.

Если теоретическое распределение не является равномерным, то необходимо задать ожидаемые (теоретические) пропорции (доли) для каждой градации (сумма долей должна быть равна 1). Для этого вместо **Expected Values: All categories equal** (Ожидаемые значения: все категории тождественны) отмечаем точкой **Expected Values: Values** (Значения). После этого вводим ожидаемую долю для наименьшей категории, затем нажимаем **Add** (Добавить), затем вводим долю для наименьшей из оставшихся категорий, и т. д. — до последней категории. Последовательность значений долей появится в нижнем окне. Нажимаем OK и получаем результаты.

Результаты (для данных примера 9.3)

А) Таблица частот (**Frequencies**)

var

	Observed N	Expected N	Residual
1.00	21	24.0	-3.0
2.00	37	24.0	13.0
3.00	29	24.0	5.0
4.00	15	24.0	-9.0
5.00	18	24.0	-6.0
Total	120		

Observed — эмпирические частоты, Expected — теоретические частоты.

В) Результаты статистической проверки (**Test statistics**):

Test Statistics

	Y
Chi-Square(a)	13.333
df	4
Asymp. Sig.	.010

a 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 24.0.

Chi-square — значение $\chi^2_э$; Asymp. Sig. — p-уровень значимости.

АНАЛИЗ ТАБЛИЦ СОПРЯЖЕННОСТИ

Анализ таблиц сопряженности применяется для решения задач, которые могут быть сформулированы следующим образом:

1. Необходимо сравнить два (или более) распределения между собой.

Например, различаются ли мужчины и женщины по распределению предпочтений пяти политических лидеров?

2. Необходимо определить связь между двумя номинативными признаками (между классификациями объектов по двум разным основаниям).

Например, связано ли соотношение предпочтений трех групп напитков (соки, лимонады, минеральные воды) с сезонностью (зима, весна, лето, осень)?

Нетрудно заметить, что эти задачи отличаются лишь словесными формулировками. Так, изучение связи между двумя номинативными переменными тождественно сравнению градаций одной номинативной переменной по распределению другой номинативной переменной.

Например, изучать сезонную зависимость предпочтений различных напитков — то же самое, что сравнивать сезоны по распределению предпочтений этих напитков. А изучать связь двух оснований классификации респондентов — по полу и по политической ориентации — то же самое, что сравнивать распределение мужчин и женщин по политической ориентации.

В подобных случаях подразумевается анализ таблиц сопряженности, в которых столбцы соответствуют сравниваемым распределениям (градациям одной номинативной переменной), а строки соответствуют градациям сравниваемых распределений (градациям другой номинативной переменной).

Формулировка проверяемой H_0: классификация объектов (людей, событий) по одному основанию не зависит от их классификации по другому основанию.

Исходные данные: определена принадлежность каждого объекта выборки к одной из градаций первой номинативной переменной и к одной из градаций второй номинативной переменной. Иными словами, две номинативные переменные измерены на выборке объектов. *Строки* таблицы сопряженности соответствуют градациям одной номинативной переменной, *столбцы* — градациям другой номинативной переменной.

Если проверка содержательной гипотезы предполагает анализ таблицы сопряженности, то принципиальным является вопрос о размерности таблицы. Будем различать два случая:

- общий случай (число градаций хотя бы одного из признаков больше 2-х),
- частный случай: таблицы сопряженности 2×2 (по две градации для каждой переменной).

Эти случаи различаются как порядком расчетов, так и особенностями интерпретации.

Число градаций больше двух

По сравнению с анализом классификации, специфика применения критерия χ^2-Пирсона (формула 9.1) к таблицам сопряженности заключается в том, что теоретические частоты рассчитываются отдельно для каждой ячейки таблицы. Таким образом, число слагаемых в формуле 9.1 равно количеству ячеек таблицы сопряженности и равно $P = k \cdot l$, где k — число строк, l — число столбцов:

$$\chi^2_э = \sum_{i=1}^{k \cdot l} \frac{(f_э - f_т)^2}{f_т}, \; df = (k-1)(l-1). \qquad (9.2)$$

Формула для расчета теоретической частоты для ячейки i-строки и j-столбца:

$$f_{ij} = \frac{f_i \cdot f_j}{N}, \qquad (9.3)$$

где f_i — сумма частот во всех ячейках i-строки; f_j — сумма частот во всех ячейках j-столбца; N — сумма частот всей таблицы сопряженности.

ПРИМЕР 9.4

Для каждого респондента репрезентативной выборки определены: а) пол; б) один из пяти предпочитаемых политических лидеров:

Эмпирические частоты		Y (политический лидер)					
		1	2	3	4	5	Всего:
X (пол)	муж.(1)	5	25	10	8	3	51
	жен. (2)	11	12	19	5	7	54
Всего:		16	37	29	13	10	105

Проверяется содержательная гипотеза о зависимости политических предпочтений от пола.

H_0: классификации объектов по двум основаниям являются независимыми (распределение объектов по полу не зависит от их распределения по предпочтениям политических лидеров).

Проверяем H_0 на уровне $\alpha = 0,05$.

Шаг 1. *Составляем таблицу сопряженности для теоретических (ожидаемых) частот* — с теми же полями, что и для таблицы эмпирических (наблюдаемых) частот. Рассчитываем значения теоретических частот для каждой ячейки этой таблицы по формуле 9.3.

для ячейки (x_1, y_1) $f_т = \frac{51 \cdot 16}{105} = 7,77$;

для ячейки (x_1, y_2) $f_т = \frac{51 \cdot 37}{105} = 17,97$;

для ячейки (x_1, y_3) $f_т = \frac{51 \cdot 29}{105} = 14,09$;

для ячейки (x_1, y_4) $f_т = \dfrac{51 \cdot 13}{105} = 6{,}31$;

для ячейки (x_1, y_5) $f_т = \dfrac{51 \cdot 10}{105} = 4{,}86$;

для ячейки (x_2, y_1) $f_т = \dfrac{54 \cdot 16}{105} = 8{,}23$;

для ячейки (x_2, y_2) $f_т = \dfrac{54 \cdot 37}{105} = 19{,}03$;

для ячейки (x_2, y_3) $f_т = \dfrac{54 \cdot 29}{105} = 14{,}91$;

для ячейки (x_2, y_4) $f_т = \dfrac{54 \cdot 13}{105} = 6{,}69$;

для ячейки (x_2, y_5) $f_т = \dfrac{54 \cdot 10}{105} = 5{,}14$.

Теоретические частоты		*Y* (политический лидер)					
		1	2	3	4	5	Всего:
X (пол)	муж. (1)	7,77	17,97	14,09	6,31	4,86	51
	жен. (2)	8,23	19,03	14,91	6,69	5,14	54
Всего:		16	37	29	13	10	105

Отметим, что суммы теоретических частот по строкам (столбцам) должны быть равны соответствующим суммам эмпирических частот.

Ш а г 2. Рассчитываем эмпирическое значение критерия χ^2-Пирсона и число степеней свободы по формуле 9.2.

$$\chi^2_э = \frac{(5-7{,}77)^2}{7{,}77} - \frac{(25-17{,}97)^2}{17{,}97} - \frac{(10-14{,}09)^2}{14{,}09} - \frac{(8-6{,}31)^2}{6{,}31} - \frac{(3-4{,}86)^2}{4{,}86} - \frac{(11-8{,}23)^2}{8{,}23} -$$

$$- \frac{(12-19{,}03)^2}{19{,}03} - \frac{(19-14{,}91)^2}{14{,}91} - \frac{(5-6{,}69)^2}{6{,}69} - \frac{(7-5{,}14)^2}{5{,}14} = 11{,}84.$$

$$df = (k-1)(l-1) = (2-1)(5-1) = 4.$$

Ш а г 3. *Определяем p-уровень по таблице критических значений χ^2-Пирсона и принимаем статистическое решение.* Для $df = 4$ наше эмпирическое значение располагается между критическими для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, p-уровень в нашем случае $p < 0{,}05$. Мы можем отклонить H_0.

Ш а г 4. *Формулируем содержательный вывод.* Обнаружена статистически значимая зависимость политических предпочтений от пола ($p < 0{,}05$).

Порядок расчетов остается тем же для любого числа градаций того и другого признака, за исключением случая таблиц сопряженности 2×2. Для упро-

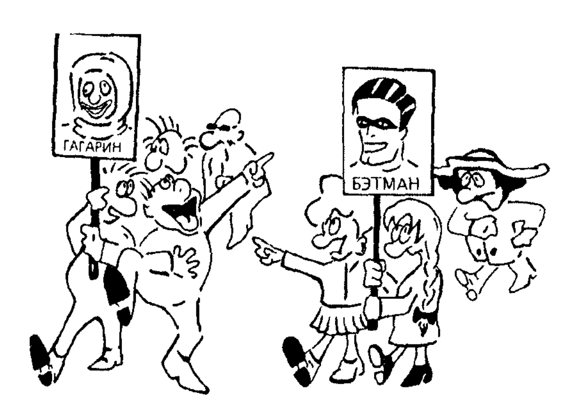
щения арифметических расчетов может быть использована формула, эквивалентная формуле 9.2:

$$\chi^2_э = N\left[\sum_{i=1}^{k}\sum_{j=1}^{l}\frac{f_{ij}^2}{f_i \times f_j} - 1\right],$$

где N — общая численность выборки; k, l — число строк и столбцов таблицы сопряженности.

Обратим внимание, что при отклонении H_0 принимается альтернативная гипотеза о связи двух оснований классификации, которая проявляется по крайней мере для одной ячейки таблицы сопряженности. Но остается неизвестным то, в отношении каких именно ячеек таблицы сопряженности связь проявляется, а в отношении каких — не проявляется. Иными словами, возникает *проблема множественных сравнений*. И для дальнейшей конкретизации результатов необходим анализ соотношения 2-х долей или таблиц сопряженности 2×2.

Исследование связи пола и предпочтений политических лидеров (см. пример 9.4) может быть продолжено. Так, может быть дополнительно проверена гипотеза о том, что лидер № 2 предпочитается чаще мужчинами, чем женщинами. Тогда необходимо сравнивать эмпирическое распределение предпочтений мужчин и женщин (25:12) с равномерным распределением (13,5:13,5) — при помощи метода сопоставления эмпирического и теоретического распределений. Может быть также проверена гипотеза о том, что лидер № 2 чаще предпочитается мужчинами, а лидер № 3 — женщинами. Тогда необходимо сопоставить два эмпирических распределения: 25:12 и 10:19 — при помощи анализа таблиц сопряженности 2×2.



Таблицы сопряженности 2×2

Существует большое разнообразие различных ситуаций, когда по результатам исследования может быть построена таблица сопряженности 2×2. Их объединяет то, что объекты (испытуемые, события) классифицированы по двум основаниям, каждое из которых представляет собой дихотомию. Важно различать два варианта такой классификации объектов:

1) *по двум различным* дихотомическим основаниям — случай независимых выборок;
2) *по одному и тому* же дихотомическому основанию дважды (например, до и после воздействия) — случай зависимых выборок.

ПРИМЕРЫ

1. Случай независимых выборок. Две группы больных известной численности получали курс лечения разными методами. Подсчитывалось число рецидивов заболевания в той и другой группе. Одна переменная — «метод лечения» (1-й, 2-й), другая — «рецидив» (есть, нет).
2. Случай зависимых выборок. Подсчитывалось число тех, кто «за», и тех, кто «против» смертной казни: до и после убедительной лекции о введении моратория на смертную казнь. Одна переменная — «до лекции» («за», «против»), другая переменная — «после лекции» («за», «против»).

Для независимых выборок применяется критерий χ^2-Пирсона, а для зависимых более адекватным является метод Мак-Нимара.

Независимые выборки

Это наиболее часто встречающаяся ситуация применения таблиц 2×2, когда одна группа объектов классифицируется по двум дихотомическим основаниям и проверяется гипотеза о связи этих двух оснований классификации.

По сравнению с другими таблицами сопряженности особенность таблиц 2×2 проявляется в трех отношениях.

1. Эти таблицы могут быть построены разными способами, но только один из них является правильным в отношении применимости критерия χ^2-Пирсона.
2. Допустима проверка направленных альтернатив. Соответственно, меняется способ определения *р*-уровня значимости.
3. В некоторых случаях при расчете χ^2-Пирсона необходимо введение поправки на непрерывность Йетса.

Рассмотрим эти особенности на примере.

ПРИМЕР 9.5

Предположим, для изучения влияния 2-х условий запоминания материала 100 испытуемых были случайным образом разделены на две группы: по 50 человек для каждого из условий. После обучения количество усвоивших этот материал в первой группе составило 24 человека, а во второй — 34 человека. Можно ли утверждать, что различия в условиях влияют на результативность обучения?

Данные примера 9.5 могут быть представлены тремя способами, но только один из них является верным.

Правильный способ представления данных примера 9.4 в таблице:

	Усвоение материала		Всего:
	есть	нет	
Условие 1	24	26	50
Условие 2	34	16	50
Всего:	58	42	100

Варианты *неправильного* представления в таблице данных примера 9.5:

	Усвоение материала	
	участвовали	усвоили
Условие 1	50	24
Условие 2	50	34

	Усвоение материала	
	наблюдаемое	ожидаемое
Условие 1	24	29
Условие 2	34	29

В последних двух случаях таблицы не содержат информации о тех, кто не усвоил материал. Поэтому уменьшаются шансы обнаружить достоверные различия, даже если они есть.

Как отмечалось, специфика применения χ^2-Пирсона в подобных случаях проявляется и в том, что *это тот случай, когда допустима проверка как ненаправленной, так и направленной статистической гипотезы*. Важность определения того, какая из этих двух гипотез проверяется, обусловлена тем, что *в отношении одних и тех же данных при проверке направленной альтернативы значение р-уровня в два раза меньше, чем при проверке ненаправленной альтернативы* (см. главу 7: Направленные и ненаправленные альтернативы).

Любые сомнения при выборе между направленной и ненаправленной статистической гипотезой решаются в пользу ненаправленной альтернативы!

Рассмотрим различия ненаправленной и направленной альтернативы в отношении данных примера 9.5. Они могли быть получены в ходе сравнения двух способов запоминания — без предварительных предположений о том, какой способ лучше. Исследователя при этом интересуют два случая (направления) отклонения H_0: а) «запоминание лучше при условии 1»; б) «запоминание лучше при условии 2». Такая проверка предполагает ненаправленную альтернативу. Соответственно, при отклонении H_0 допустим как тот, так и другой вывод. Или эти данные могли быть получены в ходе проверки предположения о том, что новый (второй) способ является более эффективным, чем традиционный (первый). Исследователя тогда будет интересовать только один исход: «запоминание лучше при условии 2». Эта проверка предполагает направленную альтернативу, а при отклонении H_0 допустим только один вывод — о превосходстве условий 2.

ПРИМЕР, КОГДА ОПРАВДАНА ПРОВЕРКА НАПРАВЛЕННОЙ ГИПОТЕЗЫ

Проверялась гипотеза о влиянии природы родства на преступность близнеца. Данные относятся к 30 преступникам мужского пола, каждый из которых имел брата близнеца. Тридцать человек были классифицированы: а) по природе родства (однояйцовые или разнояйцовые близнецы); б) по виновности или невиновности брата:

	Виновность брата:		Всего:
	виновен	не виновен	
Однояйцовый близнец	10	3	13
Разнояйцовый близнец	2	15	17
Всего:	12	18	30

(Справочник по прикладной статистике / Под ред. Э. Ллойда, У. Ледермана. М., 1989. Т. 1. С. 376).

Как указывают различные авторы, односторонний критерий χ^2-Пирсона, который применяется для ненаправленных гипотез, в данном случае «превращается» в двусторонний[1]. Таким образом, для проверки направленных гипотез *р-уровень для таблиц 2×2*, определенный по таблице для ненаправленной гипотезы (как двусторонний), *делится на 2.*

Другая особенность применения χ^2-Пирсона заключается во введении *поправки на непрерывность Йетса.* В соответствии с ней формула 9.1 для таблиц 2×2 приобретает вид:

$$\chi^2_э = \sum_{i=1}^{4} \frac{(|f_э - f_т| - 0{,}5)^2}{f_т}, \ df = 1. \qquad (9.4)$$

ПРИМЕР 9.5 (продолжение)

Предположим, данные примера 9.5 относятся к ситуации проверки содержательного предположения о большей эффективности нового метода обучения (условие 2) по сравнению с традиционным методом (условие 1).

Шаг 1. Формулируется направленная статистическая гипотеза.
Направленная H_0: При условии 2 вероятность усвоения материала не выше, чем при условии 1. В связи с тем, что объемы сравниваемых выборок не очень велики, можно принять $\alpha = 0{,}05$.

Шаг 2. Вычисляется эмпирическое значение χ^2-Пирсона с поправкой Йетса.
Теоретические частоты подсчитываем по формуле 9.3:

$$f_{ij} = \frac{f_i \cdot f_j}{N}.$$

$$f_{11} = \frac{50 \cdot 58}{100} = 29, f_{12} = \frac{50 \cdot 42}{100} = 21, f_{21} = \frac{50 \cdot 58}{100} = 29, f_{22} = \frac{50 \cdot 42}{100} = 21.$$

$f_э$	$f_т$	$\frac{(\lvert f_э - f_т\rvert - 0{,}5)^2}{f_т}$
24	29	0,698
26	21	0,964
34	29	0,698
16	21	0,964
Сумма:	100	3,325

Эмпирическое значение χ^2-Пирсона с поправкой на непрерывность $\chi^2_э = 3{,}325$.

Шаг 3. Определение *р*-уровня для направленной статистической гипотезы.
Определяем по таблице критических значений критерия χ^2-Пирсона *р*-уровень значимости. Наше эмпирическое значение располагается между критическими для $p = 0{,}1$ и $p = 0{,}05$. Следовательно, для ненаправленных гипотез в нашем случае $p < 0{,}1$. Но с учетом того, что мы проверяем направленную гипотезу, окончательное значение *р*-уровня: $p < 0{,}05$.

[1] Доказательство этого см., например: *Кендалл М., Стьюарт А.* Статистические выводы и связи. М., 1973. С. 744–745; Справочник по прикладной статистике. В 2 т. Т. 1 / Под ред. Э. Ллойда, У. Ледермана, Ю. Тюрина. М., 1989. С. 370–377.

Шаг 4. Принятие статистического решения и формулировка содержательного вывода. Статистическое решение: H_0 отклоняется. Содержательный вывод: эффективность усвоения материала в условиях обучения № 2 статистически значимо выше, чем в условиях № 1 ($\chi^2 = 3{,}325$, $df = 1$, $p < 0{,}05$).

Отметим, что при проверке ненаправленной гипотезы для тех же данных статистическое решение и, следовательно, содержательный вывод были бы другими.

χ^2-Пирсона с поправкой на непрерывность применим для анализа таблиц сопряженности 2×2, когда $N \geqslant 40$, а если ни одна из теоретических частот не меньше 5, то при $N \geqslant 20$.[1]

Если таблица сопряженности 2×2 не удовлетворяет этим требованиям (χ^2-Пирсона с поправкой на непрерывность не применим), то можно воспользоваться расчетом *точного значения р-уровня по Фишеру (Fisher's exact test — точный критерий Фишера)* — односторонним (1-sided), для направленных гипотез, или двусторонним (2-sided), для ненаправленных альтернатив. Его расчет «вручную» является трудоемким, поэтому необходимо воспользоваться компьютерной программой (SPSS, Statistica — см. конец этой главы).

Повторные измерения

Структура исходных данных соответствует ситуации, когда одна выборка объектов классифицирована на две группы дважды по одному и тому же основанию. Рассмотрим проверку гипотезы в отношении таких данных на примере.

ПРИМЕР 9.6

Исследовалось влияние убедительной лекции о введении моратория на смертную казнь. Число респондентов $N = 60$. Подсчитывалось число тех, кто «за», и тех, кто «против» смертной казни до и после лекции. Одна переменная — «до лекции» («за», «против»), другая — «после лекции» («за», «против»).

В таблице исходных данных в таких случаях каждой строке (объекту выборки) соответствуют два значения (в двух столбцах — «до», «после») одной и той же бинарной номинативной переменной («за», «против»). Таблица сопряженности для таких данных (например, построенная при помощи компьютерной программы):

		До:		Всего:
		«За»	«Против»	
После:	«За»	$a = 16$	$b = 10$	26
	«Против»	$c = 26$	$d = 8$	34
Всего:		42	18	60

Для таких данных χ^2-Пирсона с поправкой на непрерывность не применим!

[1] См. там же.

Действительно, применяя этот метод, мы будем проверять гипотезу о связи классификации ответов до лекции с классификацией ответов после лекции, а нас интересует влияние лекции («до» — «после») на распределение ответов («за» — «против»). Тем не менее, попробуем применить χ^2-Пирсона с поправкой на непрерывность к этой таблице. Получим: $\chi^2_э = 0{,}93$, $df = 1$, $p > 0{,}1$.

В подобных случаях применяется *метод Мак-Нимара.* Этот метод позволяет сопоставить долю тех, кто не обладал некоторой характеристикой (0) до воздействия, но стал обладать ею после воздействия (1), с долей тех, кто обладал этой характеристикой до воздействия (1) и перестал обладать ею после воздействия (0). Иначе говоря, метод позволяет сопоставить диагональные элементы таблицы сопряженности 2×2 (0,1 и 1,0 или 0,0 и 1,1), построенной непосредственно по дважды проведенной дихотомической классификации одной и той же выборки. Речь идет о таблице 2×2, построенной непосредственно по результатам дихотомической классификации двух зависимых выборок (одной и той же выборки — дважды):

После:	До:	
	0	1
0	*a*	*b*
1	*c*	*d*

Метод Мак-Нимара позволяет по этой таблице проверить две гипотезы: о соотношении *a* и *d* (0,1 и 1,0); о соотношении *c* и *b* (0,0 и 1,1).

Проверка гипотезы проводится по *z*-критерию по формулам для эмпирического значения[1]:

$$z_э = \frac{c-b}{\sqrt{c+b}} \text{ или } z_э = \frac{a-d}{\sqrt{a+d}}, \qquad (9.5)$$

где *c* и *b* — одна пара диагональных элементов таблицы, для проверки одной гипотезы; *a* и *d* — другая пара диагональных элементов, для проверки другой гипотезы. Для определения *p*-уровня значимости эмпирическое значение $z_э$ сравнивается с теоретическим — единичным нормальным распределением.

Ограничение на применение метода Мак-Нимара: сумма сравниваемых частот не должна быть меньше 10.

ПРИМЕР 9.6 (продолжение)

Рассмотрим применение метода Мак-Нимара на примере проверки содержательной гипотезы о влиянии лекции на мнение респондентов (данные примера 9.6).

Шаг 1. Построение таблицы 2×2.

		До:	
		«За»	«Против»
После:	«За»	*a* = 16	*b* = 10
	«Против»	*c* = 26	*d* = 8

[1] Данная реализация метода заимствована из: *Гласс Дж., Стенли Дж.* Статистические методы в педагогике и психологии. М., 1976. В программе SPSS используется критерий χ^2.

Ш а г 2. Формулировка статистической гипотезы.
Проверим H_0: $c = b$ (ненаправленная гипотеза), при $\alpha = 0{,}05$.

Отметим, что проверка гипотезы относительно других диагональных элементов (H_0: $a = d$) в данном случае не имеет смысла.

Ш а г 3. Вычисление эмпирического значения критерия.

$$z_э = \frac{c-b}{\sqrt{c+b}} = \frac{26-10}{\sqrt{26+10}} = 2{,}67 .$$

Ш а г 4. Определение p-уровня (приложение 1).
Воспользуемся таблицей единичного нормального распределения:

а) находим в таблице теоретическое значение z, *ближайшее меньшее* к абсолютному (без учета знака) эмпирическому значению $z_э$: $z_т = 2{,}65$;
б) определяем площадь под кривой справа от $z_т$: $P = 0{,}004$;
в) вычисляем p-уровень по формуле $p < 2P$: $p < 0{,}008$.

Ш а г 5. Принятие статистического решения и статистический вывод.
На уровне $\alpha = 0{,}05$ гипотеза H_0 отклоняется. Содержательный вывод: доля лиц, выступающих против смертной казни после лекции статистически значимо увеличилась ($z = 2{,}67$; $p < 0{,}008$).

Обработка на компьютере: таблицы сопряженности (кросстабуляции)

Последовательность шагов не зависит от количества градаций и зависимости выборок. Указанные обстоятельства влияют только на то, какие из результатов следует принимать во внимание.

Исходные данные: значения двух номинативных переменных (2 и более градации), с одинаковым или разным числом градаций, определены на одной выборке объектов и представлены двумя столбцами — по одному для каждой из переменных.

Выбираем: **Analyze** (Метод) > **Descriptive Statistics** (Описательные статистики) > **Crosstabs...** (Таблицы сопряженности). В открывшемся окне диалога переносим одну из переменных справа в окно Строки (**Row(s)**), другую — в окно Столбцы (**Column(s)**), нажимаем кнопку Статистики (**Statistics...**).

Решаем: Если выборки независимые (без повторных классификаций), *выбираем* χ^2, отмечая его «флажком» (**Chi-square**). Если выборки зависимые: одна и та же номинативная переменная (2 градации) измерена дважды на данной выборке, то *выбираем* метод Мак-Нимара, отмечая его «флажком» (**McNemar**). Нажимаем (**Continue**). Нажимаем OK.

Результаты

А) Сводка по обработанным объектам (Case Processing Summary) — сколько обработано (Valid), сколько пропущено (Missing), сколько всего (Total).

Б) Таблица сопряженности (Crosstabulation).

В) Таблица статистических результатов (Chi-Square Tests):

- эмпирические значения критериев (Value);
- двусторонний p-уровень для χ^2-Пирсона без поправки (с поправкой) на непрерывность (Pearson Chi-Square (Continuity Correction) — Asymp. Sig. (2-sided));
- односторонний p-уровень для направленных гипотез по Фишеру (Fisher's Exact Test — Exact Sig. (1-sided));
- двусторонний p-уровень для критерия Мак-Нимара (McNemar Test — Exact Sig. (2-sided)).

Примечание. Если обрабатываются таблицы 2×2 с независимыми классификациями, то при проверке направленных гипотез значение p-уровня для χ^2-Пирсона (Pearson Chi-Square — Asymp. Sig. (2-sided)) делится на два, либо берется односторонний p-уровень (Exact Sig. (1-sided)) для точного критерия Фишера (Fisher's Exact Test).

АНАЛИЗ ПОСЛЕДОВАТЕЛЬНОСТИ: КРИТЕРИЙ СЕРИЙ

Как следует из названия, метод применяется для анализа последовательности объектов (явлений, событий), *упорядоченных* во времени или в порядке возрастания (убывания) значений измеренного признака. Кроме того, метод требует представления последовательности в виде *бинарной переменной* — как чередования событий 0 и 1. Поэтому *исходные данные, как правило, требуют преобразования: упорядочивания (по времени или по уровню) и приведения к бинарному виду.*

Математическая идея критерия основана на подсчете числа серий в упорядоченной последовательности событий двух типов, например, 0 и 1. *Серия* — это последовательность однотипных событий, непосредственно перед и после которой произошли события другого типа. *Гипотеза* H_0 о случайном распределении событий 1 среди событий 0 может быть отклонена, если количество серий либо слишком мало, либо слишком велико.

ПРИМЕР 9.7

Предположим, было получено две последовательности успехов (1) и неудач (0) для двух игроков. Каждый из них играл 20 раз с равным количеством выигрышей ($n = 10$) и проигрышей ($m = 10$): $n + m = 20$.

Игрок № 1: 1 0 0 0 0 0 0 0 0 1 1 1 1 1 1 0 1 1 1 0 — число серий $W = 6$

Игрок № 2: 0 1 0 1 0 0 1 0 1 0 1 0 1 1 0 1 0 0 1 1 — число серий $W = 16$

В отношении первого игрока H_0 будет отклонена, если число серий слишком мало, а в отношении второго игрока — если число серий слишком велико. При отклонении H_0 для первого игрока может быть сделан вывод о том, что достоверно чаще после успеха следует успех, а после проигрыша — проигрыш, а для второго игрока, что после проигрыша достоверно чаще следует выигрыш, и наоборот.

Проблема направленности гипотезы H_0 должна решаться еще до проведения исследования. Понятно, что исследователя может интересовать любое отклонение от H_0 — как в сторону слишком малого, так и слишком большого числа серий *W*. Тогда необходима проверка ненаправленной гипотезы. Если же исследователя интересуют только малые значения *W* или только слишком большие значения *W*, то необходима проверка направленной гипотезы. Важность предварительного определения направленности гипотезы обусловлена тем, что *при одном и том же числе серий W p-уровень для направленной гипотезы будет в два раза меньше, чем для ненаправленной гипотезы*. Любые сомнения в направленности гипотезы необходимо решать в пользу выбора ненаправленной альтернативы.

Предположим, что для исследователя, получившего данные из примера 9.7, заранее не было известно, какая альтернатива будет приниматься в случае отклонения H_0. Следовательно, должна проверяться ненаправленная H_0, допускающая отклонение H_0 как в случае слишком малого, так и в случае слишком большого числа серий *W*.

Точное распределение числа серий *W* при выполнении H_0, следовательно, и *точное значение p-уровня* значимости для конкретного *W* (при конкретных значениях *m* и *n*) может быть получено с помощью комбинаторного анализа, например, при помощи компьютера.

При вычислениях на компьютере точное значение *p*-уровня может быть вычислено при выборе опции **Exact...** (Точно...) в диалоге анализа **Runs...** (Серии...) с последующим заданием метода **Monte Carlo**. Так, для примера 9.7 точные значения *p*-уровня (для ненаправленных H_0, двусторонние): для игрока № 1 $p = 0{,}035$; для игрока № 2 $p = 0{,}035$.

Если численность $m(n) < 20$, то для проверки H_0 применяются таблицы критических значений для числа серий (приложение 5).

ПРИМЕР 9.7 (продолжение)

Проверим ненаправленную H_0 в отношении двух игроков с использованием таблицы критических значений числа серий для $\alpha = 0{,}05$ (приложение 5). Для этого достаточно соотнести эмпирическое значение числа серий с табличными значениями (нижним $W_{0,025}$ и верхним $W_{0,0975}$). Если эмпирическое значение меньше или равно $W_{0,025}$ или больше или равно $W_{0,0975}$, то H_0 отклоняется.

Шаг 1. Принимаем статистические решения. Для $m = 10$, $n = 10$: $W_{0,025} = 6$; $W_{0,0975} = 16$. Для игрока № 1: $W_э = 6$, H_0 отклоняется. Для игрока № 2: $W_э = 16$, H_0 отклоняется.

Шаг 2. Формулируем содержательные выводы. Для игрока №1: достоверно чаще после успеха следует успех, а после проигрыша — проигрыш ($p < 0{,}05$). Для игрока № 2: после проигрыша достоверно чаще следует выигрыш, а после выигрыша — проигрыш.

Альтернативным способом определения *p*-уровня является применение *Z-критерия серий*, основанного на том факте, что число серий *W* при выпол-

нении H_0 распределено приблизительно нормально с известными $M_{W_э}$ и σ_W. Формула для определения *эмпирического значения Z-критерия серий*[1]:

$$Z_э = \frac{W + 0{,}5 - M_W}{\sigma_W} = \frac{W + 0{,}5 - [1 + 2nm/(n+m)]}{\sqrt{\dfrac{2nm(2nm-n-m)}{(n+m)^2(n+m-1)}}}. \qquad (9.6)$$

Ограничение на применение Z-критерия серий: $m \geqslant 20$, $n \geqslant 20$; *m* и *n* несущественно различаются. Если *m* и *n* существенно различаются, то следует воспользоваться комбинаторным методом (например, Монте Карло в программе SPSS).

ПРИМЕР 9.8

Предположим, исследуется динамика научения в игровом задании. Исследователь предполагает частые повторы проигрышей в начале и выигрышей — в конце последовательности игр (предполагается проверка направленной гипотезы). Игроком сыграно 40 партий, из них проиграно 20, выиграно 20, число серий 15. К концу последовательности игр наблюдается преобладание выигрышей. Проверим гипотезу с применением *Z*-критерия серий.

Шаг 1. Формулируем H_0: число серий соответствует случайному распределению выигрышей в последовательности проигрышей (альтернативная H_1: число серий достаточно мало, чтобы говорить о неслучайном преобладании выигрышей в конце последовательности игр). Принимаем $\alpha = 0{,}05$.

Шаг 2. Вычислим эмпирическое значение *Z*-критерия для $m = 20$; $n = 20$; $W_э = 15$:

$$M_W = 1 + 2nm/(n+m) = 21;$$

$$\sigma_W = \sqrt{\frac{2nm(2nm-n-m)}{(n+m)^2(n+m-1)}} = 3{,}12\,;\quad Z_э = \frac{15 + 0{,}5 - 21}{3{,}12} = -1{,}76\,.$$

Шаг 3. Определим *p*-уровень. Для этого воспользуемся таблицей стандартных нормальных вероятностей (приложение 1). При использовании *Z*-распределения для проверки направленной гипотезы *p*-уровень равен площади *P* под нормальной кривой справа от $+Z_э$ (слева от $-Z_э$). $Z_э = 1{,}76$ соответствует площадь $P = 0{,}039$. Следовательно, $p < 0{,}04$.

Шаг 4. Принимаем статистическое решение и формулируем содержательный вывод. Отклоняем H_0: число серий статистически значимо мало. Содержательный вывод: к концу последовательности игр статистически достоверно возрастает частота выигрышей ($p < 0{,}04$).

Отметим, что если бы проверялась ненаправленная гипотеза, то найденное значение вероятности $P = 0{,}039$ следовало бы умножить на 2: $p \leqslant 2P$. Следовательно, $p \leqslant 0{,}078$, и H_0 на уровне $\alpha = 0{,}05$ не отклоняется.

Критерий серий применим для решения двух классов задач. Помимо исследования временной последовательности событий *X* и *Y*, или динамики изменения количественного признака, метод может применяться и для провер-

[1] По Ллойду Э., Ледерману У., с. 131.

ки гипотез о различии между двумя выборками по уровню и изменчивости признака, измеренного в количественной шкале. В связи с этим применение метода требует решения проблемы преобразования исходных данных.

Проблема преобразования исходных данных. Как было отмечено, для применения метода данные необходимо представить в виде одной бинарной переменной. В зависимости от задачи исследования и вида исходных данных это может быть сделано разными способами.

1. Если изучается динамика изменчивости количественного признака, то после упорядочивания значений признака в соответствии с временной последовательностью выбирается один из способов перехода к бинарной шкале. Для метрических данных *точкой деления* (Cut point) обычно выступает среднее, а для ранговых данных — медиана. Значениям ниже точки деления присваивается 0, а значениям выше нее — 1. После такого преобразования возможно применение к переменной критерия серий.

2. Если изучается различие между выборками по уровню и (или) изменчивости количественного признака, то сначала объекты упорядочиваются по уровню выраженности изучаемой переменной. Затем объектам одной выборки присваивается 0, а объектам другой — 1. Критерий серий применяется к полученной таким образом последовательности нулей и единиц. *Преимущество критерия серий*, по сравнению с другими методами сравнения выборок, проявляется в том, что он позволяет выявить не только уровневые различия (в этом его чувствительность не очень высока), но и соотношение распределений. Например, одно распределение может быть более компактным, чем другое.

Обработка на компьютере: анализ последовательности

Исходные данные: изучаемый признак (столбец) представляет собой упорядоченную последовательность значений (по времени или по уровню выраженности). Если это последовательность во времени, то допустимы количественные значения. Если значения не количественные, то они должны представлять собой последовательность 0 и 1.

Выбираем: **Analyze** (Метод) > **Nonparametric tests...** (Непараметрические методы) > **Runs...** (Серии). В открывшемся окне диалога переносим необходимую переменную из левого в правое окно (**Test Variable List**), переменных может быть несколько.

Решаем: Выбираем точку деления (**Cut point**). Если переменная бинарная (0,1), то ставим флажок только в окошко Пользовательская и задаем «1» (**Custom: 1**). Если переменная количественная, то выбираем либо медиану (**Median**), либо среднее (**Mean**). Здесь же можем выбрать расчет точного значения *p*-уровня: нажимаем **Exact...** (Точно...) и отмечаем **Monte Carlo**. Нажимаем **Continue**. Нажимаем OK.

Результаты:

- Заданная точка деления (Test Value).
- Количество объектов ниже (выше) точки деления (Cases <(>=) Test Value).
- Общее число объектов (Total Cases).
- Число серий (Number of Runs).
- *Z*-значение (Z).
- Приблизительное значение двустороннего р-уровня (Asymp. Sig. (2-tiled)).
- Точное значение двустороннего р-уровня (Monte Carlo Sig. (2-tiled)).

П р и м е ч а н и е. Если проверяется направленная гипотеза, то значение *p*-уровня делится на 2.

Глава 10

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Корреляционный анализ — это проверка гипотез о связях между переменными с использованием коэффициентов корреляции. Наиболее распространенные коэффициенты корреляции подробно рассмотрены в главе 6. В этой главе разбираются вопросы, непосредственно касающиеся проверки гипотез с применением коэффициентов корреляции.

Коэффициент корреляции — это мера прямой или обратной пропорциональности между двумя переменными. Он чувствителен к связи только в том случае, если эта связь является монотонной — не меняет направления по мере увеличения значений одной из переменных.

Основные показатели: сила, направление и надежность (достоверность) связи. Сила связи определяется по абсолютной величине корреляции (меняется от 0 до 1). Направление связи определяется по знаку корреляции: положительный — связь прямая; отрицательный — связь обратная. Надежность связи определяется p-уровнем статистической значимости (чем меньше p-уровень, тем выше статистическая значимость, достоверность связи).

Условия применения коэффициентов корреляции:

- переменные измерены в количественной (ранговой, метрической) шкале на одной и той же выборке объектов;
- связь между переменными является монотонной.

Основная проверяемая статистическая гипотеза в отношении коэффициентов корреляции является ненаправленной и содержит утверждение о равенстве корреляции нулю в генеральной совокупности H_0: $\bar{r}_{xy} = 0$. При ее отклонении принимается альтернативная гипотеза H_1: $\bar{r}_{xy} \neq 0$ о наличии положительной (отрицательной) корреляции — в зависимости от знака выборочного (вычисленного) коэффициента корреляции.

Содержательные выводы. Если по результатам статистической проверки H_0: $\bar{r} = 0$ не отклоняется на уровне α, то содержательный вывод: *связь между x и y не обнаружена.* Если H_0: $\bar{r} = 0$ отклоняется на уровне α, то содержательный вывод: *обнаружена положительная (отрицательная) связь между x и y.*

Что влияет на p-уровень значимости корреляции? Статистическая значимость коэффициента корреляции тем выше (p-уровень меньше), чем больше его аб-

солютная величина (при одном и том же объеме выборки) и чем больше объем выборки (при одном и том же значении корреляции). При большой численности выборки даже слабые связи могут достигать статистической значимости.

Например, для одного и того же значения $r_{xy} = 0{,}200$, если $N < 90$, то $p > 0{,}05$ — корреляция статистически не значима; а если $N > 100$, то $p < 0{,}05$ — связь статистически достоверна.

Величина корреляции не всегда отражает силу связи. Соответственно, *p*-уровень значимости не всегда отражает надежность связи. Наиболее распространенные причины — «выбросы», «ложные» корреляции, нелинейные связи (см. раздел главы 6 «Величина корреляции и сила связи»).

КОРРЕЛЯЦИЯ МЕТРИЧЕСКИХ ПЕРЕМЕННЫХ

Статистическая гипотеза о связи двух метрических переменных проверяется в отношении коэффициента корреляции *r*-Пирсона, который вычисляется по формуле:

$$r_{xy} = \frac{\sum_{i=1}^{N}(x_i - M_x)(y_i - M_y)}{(N-1)\sigma_x\sigma_y}.$$

Основной (нулевой) статистической гипотезой является равенство *r*-Пирсона нулю в генеральной совокупности (H_0: $\bar{r}_{xy} = 0$). Определение *p*-уровня значимости осуществляется при помощи критерия *t*-Стьюдента:

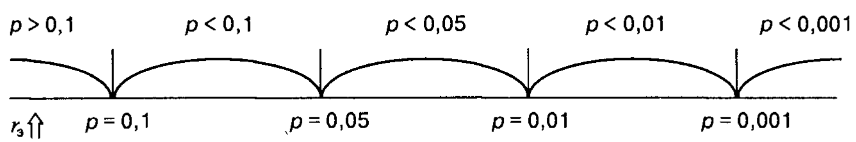
$$t_э = \frac{r_{xy}\sqrt{N-2}}{\sqrt{1-r_{xy}^2}},\ df = N-2. \qquad (10.1)$$

С целью упрощения проверки при обработке данных «вручную» обычно пользуются *таблицами критических значений* r_{xy}, которые составлены с помощью этого критерия (приложение 6). При вычислениях на компьютере статистическая программа (SPSS, Statistica) сопровождает вычисленный коэффициент корреляции более точным значением *p*-уровня.

Для статистического решения о принятии или отклонении H_0 обычно устанавливают $\alpha = 0{,}05$, а для выборок большого объема (около 100 и более) $\alpha = 0{,}01$. Если $p \leqslant \alpha$, H_0 отклоняется и делается содержательный вывод о том, что обнаружена статистически достоверная (значимая) связь между изучаемыми переменными (положительная или отрицательная — в зависимости от знака корреляции). Когда $p > \alpha$, H_0 не отклоняется, и содержательный вывод ограничен констатацией того, что связь (статистически достоверная) не обнаружена.

ПРИМЕР 10.1

На выборке $N = 20$ (учащиеся 8-го класса) были измерены два показателя интеллекта: вербального (x) и невербального (y) (см. пример 6.1). Коэффициент корреляции составил: $r_{xy} = 0{,}517$. Проверим гипотезу о связи этих показателей двумя способами. Подставив величины $N = 20$ и $r_{xy} = 0{,}517$ в формулу 10.1, получаем: $t_э = 2{,}562$; $df = 18$. По таблице критических значений t-Стьюдента (приложение 2) для $df = 18$ видим, что эмпирическое значение находится между критическими значениями для $p = 0{,}05$ и $p = 0{,}01$.



Следовательно, для нашего случая $p < 0{,}05$. Тот же результат мы получим, минуя вычисление t-Стьюдента, воспользовавшись таблицей критических значений коэффициента корреляции r-Пирсона (приложение 6): в строке, соответствующей $N = 20$, видим, что эмпирическое значение корреляции находится между критическими значениями для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$. (При расчете на компьютере значение коэффициента корреляции будет сопровождаться точным значением p-уровня, для данного случая: $p = 0{,}019$.)

Статистическое решение: H_0: $\bar{r}_{xy} = 0$ отклоняется для $\alpha = 0{,}05$. *Содержательный вывод*: обнаружена статистически достоверная положительная связь вербального и невербального интеллекта для учащихся 8-го класса ($r_{xy} = 0{,}517$, $N = 20$, $p < 0{,}05$).

Замечания к применению метрических коэффициентов корреляции. *Если связь (статистически достоверная) не обнаружена,* но есть основания полагать, что связь на самом деле есть, то следует проверить возможные причины недостоверности связи.

1. Нелинейность связи: просмотреть график двумерного рассеивания. Если связь нелинейная, но монотонная, перейти к ранговым корреляциям. Если связь не монотонная, то делить выборку на части, в которых связь монотонная, и вычислить корреляции отдельно для каждой части выборки, или делить выборку на контрастные группы и далее сравнивать их по уровню выраженности признака.

2. Наличие выбросов и выраженная асимметрия распределения одного или обоих признаков. Просмотреть гистограммы распределения частот того и другого признака. При наличии выбросов или асимметрии исключить выбросы или перейти к ранговым корреляциям.

3. Неоднородность выборки: просмотреть график двумерного рассеивания. Попытаться разделить выборку на части, в которых связь может иметь разные направления.



Если связь не обнаружена, но есть основания полагать, что связь на самом деле есть...

Если связь статистически достоверна, то прежде, чем делать содержательный вывод, следует исключить возможность «ложной» корреляции.

1. Связь обусловлена выбросами: просмотреть график двумерного рассеивания. При наличии выбросов перейти к ранговым корреляциям или исключить выбросы.

2. Связь обусловлена влиянием третьей переменной: просмотреть график двумерного рассеивания на предмет наличия содержательно интерпретируемого деления выборки на группы, для которых согласованно меняются средние двух переменных. Если подобное явление возможно, необходимо вычислить корреляцию не только для всей выборки, но и для каждой группы в отдельности. Если «третья» переменная метрическая — вычислить частную корреляцию.

ЧАСТНАЯ КОРРЕЛЯЦИЯ

Если изучается связь между тремя метрическими переменными, то возможна проверка предположения о том, что связь между двумя переменными X и Y не зависит от влияния третьей переменной — Z. Для этого можно вычислить коэффициент частной корреляции r_{xy-z}:

$$r_{xy-z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1-r_{xz}^2)(1-r_{yz}^2)}}.$$

Напомним, что коэффициент r_{xy-z} тем больше по абсолютной величине (ближе к r_{xy}), чем меньше связь между X и Y обусловлена влиянием Z. Коэффициент r_{xy-z} близок к 0, если связь между X и Y близка к 0 при любом фиксированном значении Z, то есть связь между X и Y обусловлена влиянием Z.

Основной (нулевой) статистической гипотезой является равенство частной корреляции нулю в генеральной совокупности (H_0: $\bar{r}_{xy-z} = 0$). Определение p-уровня значимости осуществляется при помощи критерия t-Стьюдента:

$$t_э = \frac{r_{xy-z}\sqrt{N-3}}{\sqrt{1-r_{xy-z}^2}},\ df = N-3, \qquad (10.2)$$

Если $p \leqslant \alpha$, H_0 отклоняется и делается содержательный вывод о том, что обнаружена статистически достоверная связь x и y при фиксированных значениях z, то есть связь между x и y не зависит от влияния z. Когда $p > \alpha$, H_0 не отклоняется, и содержательный вывод ограничен констатацией того, что связь (статистически достоверная) между x и y при фиксированных значениях z не обнаружена.

ПРОВЕРКА ГИПОТЕЗ О РАЗЛИЧИИ КОРРЕЛЯЦИЙ

Задача сравнения корреляций имеет два варианта решения: а) *для независимых выборок* — когда необходимо сравнить два коэффициента корреляции, полученных на разных выборках между одними и теми же переменными; б) *для зависимых выборок* — когда необходимо сравнить корреляцию переменных X и Y с корреляцией переменных X и Z, при условии, что все три переменные измерены на одной и той же выборке[1].

Сравнение корреляций для независимых выборок

По результатам сравнения корреляций в данном случае можно делать вывод о различии корреляции признаков X и Y в двух сравниваемых совокупностях. Проверяемая H_0 содержит утверждение о равенстве корреляций в генеральной совокупности.

ПРИМЕР 10.2

В одном исследовании сравнивалась связь интеллекта и среднего балла отметок учащихся 6-х классов и учащихся 11-х классов. Для 50 учащихся 6-х классов корреляция составила $r_1 = 0,63$ ($p < 0,001$), а для 60 учащихся 11-х классов — $r_2 = 0,31$ ($p < 0,05$). Можно ли на основании этих данных утверждать, что в 11-х классах связь отметок с интеллектом слабее, чем в 6-х классах?

Задача статистической проверки подобных предположений решается при помощи Z-преобразования Фишера коэффициентов корреляции и последующего применения Z-критерия. Z-преобразование Фишера — это пересчет коэффициентов корреляции r по формуле:

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r}\cdot \qquad (10.3)$$

Для облегчения пересчета можно воспользоваться функцией «ФИШЕР» в программе Excel либо таблицей, составленной с ее помощью (приложение 7).

Эмпирическое значение Z-критерия для определения p-уровня значимости различия корреляций вычисляется по формуле:

$$Z_э = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{N_1-3}+\frac{1}{N_2-3}}}, \qquad (10.4)$$

[1] Методы этого раздела заимствованы из: *Гласс Дж., Стенли Дж.* Статистические методы в педагогике и психологии. М., 1977. С. 283–286.

где Z_1 и Z_2 — Z-преобразованные значения сравниваемых корреляций, N_1 и N_2 — соответствующие объемы выборок. Уровень значимости определяется по формуле $p < 2P$, где P — площадь справа от $Z_э$ под кривой нормального распределения.

ПРИМЕР 10.2 (продолжение)

Проверим гипотезу о различии коэффициентов корреляции ($\alpha = 0{,}05$).

Ш а г 1. Производим Z-преобразование Фишера в отношении сравниваемых корреляций, воспользовавшись таблицей из приложения 7:

$$Z_1 = 0{,}741;\ Z_2 = 0{,}321.$$

Ш а г 2. Вычислим эмпирическое значение Z-критерия по формуле 10.4:

$$Z_э = \frac{0{,}741 - 0{,}321}{\sqrt{\dfrac{1}{50-3} + \dfrac{1}{60-3}}} = 2{,}136\ .$$

Ш а г 3. Определим p-уровень значимости. По таблице стандартных нормальных вероятностей (приложение 1) определяем площадь справа от табличного z, ближайшего *меньшего* $Z_э$. Справа от $z = 2{,}13$: $P = 0{,}0166$. Уровень значимости определяется по формуле $p < 2P$. Следовательно, $p < 0{,}033$.

Ш а г 4. Принимаем статистическое решение и формулируем содержательный вывод. Статистическое решение: отклоняем H_0 (о равенстве корреляций в генеральной совокупности). Содержательный вывод: в 11-х классах связь отметок с интеллектом статистически значимо ниже, чем в 6-х классах ($p < 0{,}033$).

Отметим, что одна и та же разность между корреляциями будет иметь более высокую статистическую значимость при больших значениях корреляции и меньшую — при более слабых корреляциях. Так, уменьшение значений корреляций всего на 0,1 в примере 10.2 привело бы к $p > 0{,}05$.

Сравнение корреляций для зависимых выборок

В данном случае предполагается сравнение корреляции X и Y с корреляцией X и Z при условии, что все три признака измерены на одной и той же выборке. Проверяемая H_0 содержит утверждение о равенстве соответствующих корреляций.

ПРИМЕР 10.3

Сравнивалась прогностическая эффективность двух шкал вступительного теста в отношении предсказания среднего балла отметок студентов 2 курса. На выборке в 95 студентов корреляция результатов тестирования и среднего балла отметок составила: для первой шкалы: $r_1 = 0{,}60$; для второй шкалы: $r_2 = 0{,}46$; корреляция результатов двух тестов: $r_{12} = 0{,}70$. Можно ли утверждать, что прогностическая ценность первой шкалы достоверно выше, чем второй?

Для статистической проверки подобных гипотез применяется Z-критерий, эмпирическое значение которого вычисляется по формуле:

$$Z_э = \frac{(r_{xy} - r_{xz})\sqrt{N}}{\sqrt{(1 - r_{xy}^2)^2 + (1 - r_{xz}^2)^2 - 2r_{yz}^3 - (2r_{yz} - r_{xy}r_{xz})(1 - r_{xy}^2 - r_{xz}^2 - r_{yz}^2)}}. \qquad (10.5)$$

ПРИМЕР 10.3 (продолжение)

Проверим гипотезу о различии коэффициентов корреляции ($\alpha = 0,05$).

Шаг 1. Вычислим эмпирическое значение Z-критерия по формуле 10.5: $Z_э = 2,119$.

Шаг 2. Определим p-уровень значимости. По таблице стандартных нормальных вероятностей (приложение 1) определяем площадь справа от табличного z, ближайшего *меньшего* $Z_э$. Справа от $z = 2,11$: $P = 0,0174$. Уровень значимости определяется по формуле $p < 2P$. Следовательно, $p < 0,035$.

Шаг 3. Принимаем статистическое решение и формулируем содержательный вывод. Статистическое решение: отклоняем H_0 (о равенстве корреляций в генеральной совокупности). Содержательный вывод: корреляция второй шкалы теста статистически достоверно ниже корреляции первой шкалы со средним баллом отметок студентов 2-го курса ($p < 0,05$) — прогностическая ценность первой шкалы выше, чем второй шкалы.

Отметим, что для решения такой задачи можно было бы рассматривать выборки как независимые и применять соответствующий метод сравнения корреляций — по формулам 10.3 и 10.4. Но чувствительность (мощность) такой проверки была бы гораздо ниже. В частности, применяя к данным примера 10.3 предыдущий метод, мы получим $p = 0,18$, что приводит к принятию H_0.

КОРРЕЛЯЦИЯ РАНГОВЫХ ПЕРЕМЕННЫХ

Если к количественным данным неприменим коэффициент корреляции r-Пирсона, то для проверки гипотезы о связи двух переменных *после предварительного ранжирования* могут быть применены корреляции r-Спирмена или τ-Кендалла.

***r*-Спирмена.** Этот коэффициент корреляции вычисляется либо путем применения формулы r-Пирсона к предварительно ранжированным двум переменным, либо, при отсутствии повторяющихся рангов, по упрощенной формуле:

$$r_s = 1 - \frac{6\sum_i d_i^2}{N(N^2 - 1)}.$$

Поскольку этот коэффициент — аналог r-Пирсона, то и применение r-Спирмена для проверки гипотез аналогично применению r-Пирсона, изложенному ранее[1].

Преимущество r-Спирмена по сравнению с r-Пирсона — в большей чувствительности к связи в случае:

- существенного отклонения распределения хотя бы одной переменной от нормального вида (асимметрия, выбросы);
- криволинейной (монотонной) связи.

Недостаток r-Спирмена по сравнению с r-Пирсона — в меньшей чувствительности к связи в случае несущественного отклонения распределения обеих переменных от нормального вида.

Частная корреляция и сравнение корреляций применимы и к r-Спирмена.

τ-Кендалла. Применяется к предварительно ранжированным данным как альтернатива r-Спирмена. τ-Кендалла, как отмечалось в главе 6, имеет более выгодную, вероятностную интерпретацию. Общая формула для вычисления τ-Кендалла, вне зависимости от наличия или отсутствия повторяющихся рангов (связей):

$$\tau_b = \frac{P-Q}{\sqrt{[N(N-1)/2]-K_x}\sqrt{[N(N-1)/2]-K_y}},$$

где P — число совпадений, Q — число инверсий, K_x и K_y — поправки на связи в рангах (см. главу 6: Проблема связанных (одинаковых) рангов). Если связей в рангах нет, то знаменатель формулы равен $P + Q = N(N-1)/2$.

Поскольку природа τ-Кендалла иная, чем у r-Спирмена и r-Пирсона, то p-уровень определяется по-другому: применяется z-критерий и единичное нормальное распределение. Эмпирическое значение вычисляется по формуле:

$$z_э = \frac{|P-Q|-1}{\sqrt{N(N-1)(2N+5)/18}}. \qquad (10.6)$$

При вычислениях «вручную» p-уровень определяется по следующему алгоритму:

а) вычисляется эмпирическое значение $z_э$;

б) по таблице «Стандартные нормальные вероятности» (приложение 1) определяется теоретическое значение z, *ближайшее меньшее к эмпирическому значению* $z_э$;

в) определяется площадь P под кривой справа от $z_т$;

г) вычисляется p-уровень по формуле $p < 2P$.

Проверяемая статистическая гипотеза, порядок принятия статистического решения и формулировка содержательного вывода те же, что и для случая r-Пирсона или r-Спирмена.

[1] В некоторых источниках по непонятным причинам для r-Пирсона и r-Спирмена приводят разные таблицы критических значений. В компьютерных программах (SPSS, STATISTICA) уровни значимости для одинаковых r-Пирсона и r-Спирмена всегда совпадают.

При вычислениях на компьютере статистическая программа (SPSS, Statistica) сопровождает вычисленный коэффициент корреляции более точным значением p-уровня.

ПРИМЕР 10.4

Предположим, для каждого из 12 учащихся одного класса известно время решения тестовой арифметической задачи в секундах (X) и средний балл отметок по математике за последнюю четверть (Y). При подсчете τ-Кендалла были получены следующие результаты: $P = 18$; $Q = 48$; $\tau = -0{,}455$. Проверим гипотезу о связи времени решения тестовой задачи и среднего балла отметок по математике.

Шаг 1. Вычисляем эмпирическое значение критерия:

$$z_э = \frac{|18-48|-1}{\sqrt{12(12-1)(2\cdot 12+5)/18}} = 1{,}989 .$$

Шаг 2. По таблице «Стандартные нормальные вероятности» (приложение 1) находим ближайшее меньшее, чем $z_э$, теоретическое значение $z_т$ и площадь справа от этого $z_т$: $z_т = 1{,}98$; площадь справа $P = 0{,}024$.

Шаг 3. Вычисляем p-уровень по формуле $p < 2P$; $p < 0{,}048$.

Шаг 4. Принимаем статистическое решение. Нулевая гипотеза об отсутствии связи в генеральной совокупности отклоняется на уровне $\alpha = 0{,}05$.

Шаг 5. Формулируем содержательный вывод. Обнаружена отрицательная связь между временем решения тестовой арифметической задачи и средним баллом отметок по математике за последнюю четверть ($\tau = -0{,}455$; $N = 12$; $p < 0{,}048$).
Величина корреляции показывает, что при сравнении испытуемых друг с другом более высокий средний балл будет сочетаться с меньшим временем решения задач чаще, чем в 70% случаях, так как вероятность инверсий $P(q) = (1-\tau)/2 = (1+0{,}455)/2 = 0{,}728$.

(Отметим, что при вычислении τ-Кендалла по этим данным на компьютере были получены следующие результаты: $\tau = -0{,}455$; $p = 0{,}040$.)

Сравнение r-Спирмена и τ-Кендалла. Интерпретация r-Спирмена аналогична интерпретации r-Пирсона. Квадрат и того, и другого коэффициента корреляции (коэффициент детерминации) показывает долю дисперсии одной переменной, которая может быть объяснена влиянием другой переменной. τ-Кендалла имеет другую интерпретацию: это разность вероятностей совпадений и инверсий в рангах. Кроме того, по величине τ-Кендалла можно судить о вероятности совпадений $P(p) = (1+\tau)/2$ или инверсий $P(q) = (1-\tau)/2$.

Для одних и тех же данных величина r-Спирмена всегда больше, чем τ-Кендалла, исключая крайние значения 0 и 1. Это отражает тот факт, что τ-Кендалла зависит от силы связи линейно, а r-Спирмена — не линейно. В то же время для одних и тех же данных *p-уровень τ-Кендалла и r-Спирмена примерно одинаков*, а иногда τ-Кендалла имеет преимущество в уровне значимости.

Замечания к применению. *Если связь (статистически достоверная) не обнаружена,* но есть основания полагать, что связь на самом деле есть, то следует

сначала перейти от r-Спирмена к τ-Кендалла (или наоборот), а затем проверить другие возможные причины недостоверности связи.

1. Нелинейность связи: просмотреть график двумерного рассеивания. Если связь не монотонная, то делить выборку на части, в которых связь монотонная, или делить выборку на контрастные группы и далее сравнивать их по уровню выраженности признака.
2. Неоднородность выборки: просмотреть график двумерного рассеивания. Попытаться разделить выборку на части, в которых связь может иметь разные направления.

Если связь статистически достоверна, то прежде, чем делать содержательный вывод, следует исключить возможность наличия «ложной» корреляции, как следствия влияния третьей переменной (см. Замечания к применению метрических коэффициентов корреляции).

АНАЛИЗ КОРРЕЛЯЦИОННЫХ МАТРИЦ

Корреляционная матрица. Часто корреляционный анализ включает в себя изучение связей не двух, а множества переменных, измеренных в количественной шкале на одной выборке. В этом случае вычисляются корреляции для каждой пары из этого множества переменных. Вычисления обычно проводятся на компьютере, а результатом является корреляционная матрица.

Корреляционная матрица (*Correlation Matrix*) — это результат вычисления корреляций одного типа для каждой пары из множества P переменных, измеренных в количественной шкале на одной выборке.

ПРИМЕР

Предположим, изучаются связи между 5 переменными ($v1$, $v2$, ..., $v5$; $P = 5$), измеренными на выборке численностью $N = 30$ человек. Ниже приведена таблица исходных данных и корреляционная матрица.

Исходные данные:

№	v1	v2	v3	v4	v5
1	10	23	4	111	56
2	13	26	6	98	52
3	8	12	2	105	58
4	9	25	7	100	49
5	11	16	3	101	65
...	...	...	...	...	...
30	7	19	6	94	41

Корреляционная матрица:

	v1	v2	v3	v4	v5
v1	1	0,52	−0,11	−0,29	−0,38
v2	0,52	1	0,28	0,32	−0,34
v3	−0,11	0,28	1	0,48	0,42
v4	−0,29	0,32	0,48	1	0,38
v5	−0,38	−0,34	0,42	0,38	1

Нетрудно заметить, что корреляционная матрица является квадратной, симметричной относительно главной диагонали (так как $r_{ij} = r_{ji}$), с единицами на главной диагонали (так как $r_{ii} = r_{jj} = 1$).

Корреляционная матрица является *квадратной*: число строк и столбцов равно числу переменных. Она *симметрична* относительно главной диагонали, так как корреляция x с y равна корреляции y с x. На ее главной диагонали располагаются единицы, так как корреляция признака с самим собой равна единице. Следовательно, анализу подлежат не все элементы корреляционной матрицы, а те, которые находятся выше или ниже главной диагонали.

Количество коэффициентов корреляции, подлежащих анализу при изучении связей P признаков определяется формулой: $P(P-1)/2$. В приведенном выше примере количество таких коэффициентов корреляции $5(5-1)/2 = 10$.

Основная задача анализа корреляционной матрицы — выявление структуры взаимосвязей множества признаков. При этом возможен визуальный анализ *корреляционных плеяд* — графического изображения *структуры статистически значимых связей*, если таких связей не очень много (до 10–15). Другой способ — применение многомерных методов: множественного регрессионного, факторного или кластерного анализа (см. раздел «Многомерные методы...»). Применяя факторный или кластерный анализ, можно выделить группировки переменных, которые теснее связаны друг с другом, чем с другими переменными. Весьма эффективно и сочетание этих методов, например, если признаков много и они не однородны.

Сравнение корреляций — дополнительная задача анализа корреляционной матрицы, имеющая два варианта. Если необходимо сравнение корреляций в одной из строк корреляционной матрицы (для одной из переменных), применяется метод сравнения для зависимых выборок (с. 148–149). При сравнении одноименных корреляций, вычисленных для разных выборок, применяется метод сравнения для независимых выборок (с. 147–148).

Методы сравнения корреляций *в диагоналях* корреляционной матрицы (для оценки стационарности случайного процесса) и сравнения *нескольких* корреляционных матриц, полученных для разных выборок (на предмет их однородности), являются трудоемкими и выходят за рамки данной книги. Познакомиться с этими методами можно по книге Г. В. Суходольского[1].

Проблема статистической значимости корреляций. Проблема заключается в том, что процедура статистической проверки гипотезы предполагает *однократное* испытание, проведенное на одной выборке. Если один и тот же метод применяется *многократно*, пусть даже и в отношении различных переменных, то увеличивается вероятность получить результат чисто случайно. В общем случае, если мы повторяем один и тот же метод проверки гипотезы k раз в отношении разных переменных или выборок, то при установленной величине α мы гарантированно получим подтверждение гипотезы в $\alpha \times k$ числе случаев.

Предположим, анализируется корреляционная матрица для 15 переменных, то есть вычислено $15(15-1)/2 = 105$ коэффициентов корреляции. Для проверки гипотез установлен уровень $\alpha = 0,05$. Проверяя гипотезу 105 раз, мы пять раз (!) получим ее подтверждение независимо от того, существует ли связь на самом деле. Зная это и

[1] *Суходольский Г. В.* Основы математической статистики для психологов. СПб., 1998. С. 299–302.

получив, скажем, 15 «статистически достоверных» коэффициентов корреляции, сможем ли мы сказать, какие из них получены случайно, а какие — отражают реальную связь?

Строго говоря, для принятия статистического решения необходимо уменьшить уровень α во столько раз, сколько гипотез проверяется. Но вряд ли это целесообразно, так как непредсказуемым образом увеличивается вероятность проигнорировать реально существующую связь (допустить ошибку II рода).

Одна только корреляционная матрица не является достаточным основанием для статистических выводов относительно входящих в нее отдельных коэффициентов корреляций!

Можно указать лишь один действительно убедительный способ решения этой проблемы: разделить выборку случайным образом на две части и принимать во внимание только те корреляции, которые статистически значимы в обеих частях выборки. Альтернативой может являться использование многомерных методов (факторного, кластерного или множественного регрессионного анализа) — для выделения и последующей интерпретации групп статистически значимо связанных переменных.

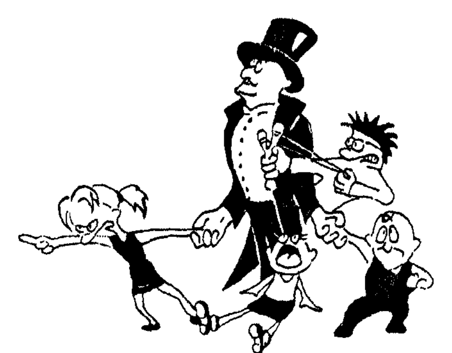
Проблема пропущенных значений. Если в данных есть пропущенные значения, то возможны два варианта расчета корреляционной матрицы: а) построчное удаление значений (**Exclude cases listwise**); б) попарное удаление значений (**Exclude cases pairwise**). При *построчном удалении* наблюдений с пропусками удаляется вся строка для объекта (испытуемого), который имеет хотя бы одно пропущенное значение по одной из переменных. Этот способ приводит к «правильной» корреляционной матрице в том смысле, что все коэффициенты вычислены по одному и тому же множеству объектов. Однако если пропущенные значения распределены случайным образом в переменных, то данный метод может привести к тому, что в рассматриваемом множестве данных не останется ни одного объекта (в каждой строке встретится, по крайней мере, одно пропущенное значение). Чтобы избежать подобной ситуации, используют другой способ, называемый *попарным удалением.* В этом способе учитываются только пропуски в каждой выбранной паре столбцов-переменных и игнорируются пропуски в других переменных. Корреляция для пары переменных вычисляется по тем объектам, где нет пропусков. Во многих ситуациях, особенно когда число пропусков относительно мало, скажем 10%, и пропуски распределены достаточно хаотично, этот метод не приводит к серьезным ошибкам. Однако иногда это не так. Например, в систематическом смещении (сдвиге) оценки может «скрываться» систематическое расположение пропусков, являющееся причиной различия коэффициентов корреляции, построенных по разным подмножествам (например — для разных подгрупп объектов). Другая проблема, связанная с корреляционной матрицей, вычисленной при *попарном* удалении пропусков, возникает при использовании этой матрицы в других видах анализа (например, в множественном регрессионном или факторном анализе). В них предполагается, что используется «правильная» корреляционная матрица с определенным уровнем состоятельности и «соответствия» различных коэффициентов. Использование матрицы с «плохими» (смещенными) оценками при-

водит к тому, что программа либо не в состоянии анализировать такую матрицу, либо результаты будут ошибочными. Поэтому, если применяется попарный метод исключения пропущенных данных, необходимо проверить, имеются или нет систематические закономерности в распределении пропусков.

Если попарное исключение пропущенных данных не приводит к какому-либо систематическому сдвигу средних значений и дисперсий (стандартных отклонений), то эти статистики будут похожи на аналогичные показатели, вычисленные при построчном способе удаления пропусков. Если наблюдается значительное различие, то есть основание предполагать наличие сдвига в оценках. Например, если среднее (или стандартное отклонение) значений переменной *A*, которое использовалось при вычислении ее корреляции с переменной *B*, намного меньше среднего (или стандартного отклонения) тех же значений переменной *A*, которые использовались при вычислении ее корреляции с переменной *C*, то имеются все основания ожидать, что эти две корреляции (*A*–*B* и *A*–*C*) основаны на разных подмножествах данных. В корреляциях будет сдвиг, вызванный неслучайным расположением пропусков в значениях переменных.

Анализ корреляционных плеяд. После решения проблемы статистической значимости элементов корреляционной матрицы статистически значимые корреляции можно представить графически в виде корреляционной плеяды или плеяд. *Корреляционная плеяда* — это фигура, состоящая из вершин и соединяющих их линий. Вершины соответствуют признакам и обозначаются обычно цифрами — номерами переменных. Линии соответствуют статистически достоверным связям и графически выражают знак, а иногда — и *p*-уровень значимости связи.

Корреляционная плеяда может отражать *все* статистически значимые связи корреляционной матрицы (иногда называется *корреляционным графом*) или только их содержательно выделенную часть (например, соответствующую одному фактору по результатам факторного анализа).



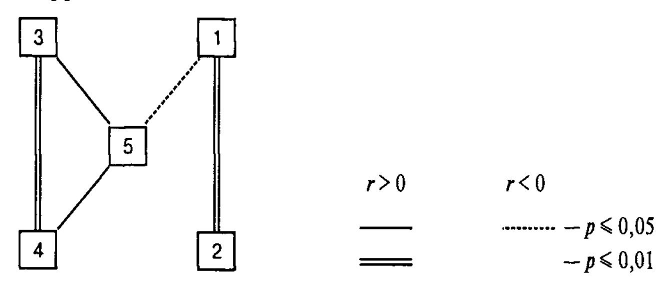
Корреляционный граф и его родственные связи, достоверность которых была установлена в судебном порядке

ПРИМЕР ПОСТРОЕНИЯ КОРРЕЛЯЦИОННОЙ ПЛЕЯДЫ

Корреляционная матрица:

	v1	v2	v3	v4	v5
v1	1	0,52	–0,11	–0,29	–0,38
v2	0,52	1	0,28	0,32	–0,34
v3	–0,11	0,28	1	0,48	0,42
v4	–0,29	0,32	0,48	1	0,38
v5	–0,38	–0,34	0,42	0,38	1
	2	1	2	2	3

Корреляционная плеяда:



Построение корреляционной плеяды начинают с выделения в корреляционной матрице статистически значимых корреляций (иногда — разным цветом в зависимости от p-уровня значимости). Затем для строк (столбцов) матрицы, содержащих статистически значимые корреляции, подсчитывается их количество. Построение плеяды начинают с переменной, имеющей наибольшее число значимых связей, постепенно добавляя в рисунок другие переменные — по мере убывания числа связей и связывая их линиями, соответствующими связям между ними.

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Графики двумерного рассеивания. Выбираем **Graphs... > Scatter...-Simple**. Нажимаем **Define**. В появляющемся окне назначаем осям переменные: выделяем слева одну переменную, нажимаем ▻ напротив «**X Axis**» (Ось X), выделяем другую переменную, нажимаем ▻ напротив «**Y Axis**». Нажимаем ОК. Получаем график рассеивания назначенных переменных.

Вычисление симметричной корреляционной матрицы. (По умолчанию SPSS вычисляет полную корреляционную матрицу.)

Выбираем **Analyze > Correlate > Bivariate...** В открывшемся окне диалога выделяем интересующие переменные в левой части и переносим их в правую часть при помощи кнопки ▻ (переменных должно быть как минимум две).

По умолчанию стоит флажок **Pearson** (корреляция r-Пирсона). Если интересует корреляция *r-Спирмена* или *τ-Кендалла*, необходимо поставить соответствующие флажки внизу.

Если в данных есть пропуски, то по умолчанию программа учтет их путем *попарного удаления* (**exclude cases pairwise**). Если необходимо учесть их путем *построчного удаления* (объектов с пропусками), то нажимаем **Options... > (Exclude cases listwise) > Continue...**

Нажимаем ОК. В появившейся таблице строки и столбцы соответствуют выделенным ранее переменным. В ячейке на пересечении строки и столбца, соответствующих интересующим нас переменным, видим три числа: верхнее

соответствует коэффициенту корреляции, нижнее — численности выборки *N*, среднее — *p*-уровню значимости для ненаправленных альтернатив (Sig. (2-tailed)).

Вычисление несимметричной корреляционной матрицы. Если есть необходимость вычислить корреляции не всех, а только двух групп переменных, то необходимо создание командного файла (Syntax). Например, есть 5 переменных с именами: *v*1, *v*2, *v*3, *v*4, *v*5. Задача — вычислить корреляции *v*1 с остальными переменными из этого набора, обрабатывая пропуски путем попарного удаления.

- Выбираем **File > New > Syntax.** В открывшемся окне набираем текст:
  ```
  correlations variables v1 with v2 v3 v4 v5.
  ```
  (Количество переменных до и после слова `with` — не ограничено).
- Если необходима обработка пропусков путем построчного удаления, то:
  ```
  correlations variables v1 with v2 v3 v4 v5
  /missing listwise.
  ```
- Если надо вычислить корреляцию *r*-Спирмена (с попарным удалением), то:
  ```
  nonpar corr v1 with v2 v3 v4 v5.
  ```
- Для вычисления корреляций τ-Кендалла добавляем к первой — вторую строку:
  ```
  nonpar corr v1 with v2 v3 v4 v5
  /print kendall.
  ```
- Для вычисления и *r*-Спирмена, и τ-Кендалла с построчным удалением:
  ```
  nonpar corr v1 with v2 v3 v4 v5
  /missing listwise
  /print both.
  ```

Заметьте, что вся команда обязательно должна заканчиваться точкой.

Для выполнения команды нажимаем **Run > All**. Программа выдаст результат — таблицу корреляций переменных. Строки будут соответствовать именам переменных, указанных в команде до слова `with`, а столбцы — именам переменных, указанных после слова `with`.

Вычисление частной корреляции. Выбираем **Analyze > Correlate > Partial...** В открывшемся окне диалога переносим интересующие переменные из левой части в правую верхнюю (**Variables:**) при помощи верхней кнопки ▻ (переменных должно быть как минимум две). Затем при помощи нижней кнопки ▻ из правой части в левую нижнюю часть (**Controlling for:**) переносим переменную, значения которой хотим фиксировать. Нажимаем ОК. Получаем таблицу, аналогичную таблице парных корреляций, но верхнее число в каждой ячейке — значение частной корреляции соответствующих двух переменных при фиксированном значении указанной третьей переменной. Нижнее число — *p*-уровень значимости, а посередине — число степеней свободы.

Глава 11

ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ ДВУХ ВЫБОРОК

Сравнение двух выборок по признаку, измеренному в метрической шкале, обычно предполагает *сравнение средних значений с использованием параметрического критерия t-Стьюдента*. Следует различать три ситуации по соотношению выборок между собой: случай *независимых* и *зависимых* выборок (измерений признака) и дополнительно — случай сравнения одного среднего значения с заданной величиной (критерий *t*-Стьюдента *для одной выборки*).

К параметрическим методам относится и *сравнение дисперсий двух выборок по критерию F-Фишера*. Иногда этот метод приводит к ценным содержательным выводам, а в случае сравнения средних для независимых выборок сравнение дисперсий является *обязательной* процедурой.

При сравнении средних или дисперсии двух выборок проверяется *ненаправленная статистическая гипотеза* о равенстве средних (дисперсий) в генеральной совокупности. Соответственно, при ее отклонении допустимо принятие двусторонней альтернативы о конкретном направлении различий в соответствии с соотношением выборочных средних (дисперсий). Для принятия статистического решения в таких случаях применяются двусторонние критерии и, соответственно, критические значения для проверки ненаправленных альтернатив.

СРАВНЕНИЕ ДИСПЕРСИЙ

Метод позволяет проверить гипотезу о том, что дисперсии двух генеральных совокупностей, из которых извлечены сравниваемые выборки, отличаются друг от друга. Проверяемая статистическая гипотеза H_0: $\sigma_1^2 = \sigma_2^2$. При ее отклонении принимается альтернативная гипотеза о том, что одна дисперсия больше другой.

Исходные предположения: две выборки извлекаются случайно из разных генеральных совокупностей с нормальным распределением изучаемого признака.

Структура исходных данных: изучаемый признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из двух сравниваемых выборок.

ПРИМЕР

При сравнении мужчин (1) и женщин (2) по уровню тревожности:

№	X (пол)	Y (тревожность)
1	1	10
2	2	9
3	2	3
4	1	8
...	...	...
N	1	6

Ограничения: распределения признака и в той, и в другой выборке существенно не отличаются от нормального.

Альтернатива методу: критерий Ливена (Levene's Test), применение которого не требует проверки предположения о нормальности (используется в программе SPSS).

Формула для эмпирического значения критерия F-Фишера:

$$F_э = \frac{\sigma_1^2}{\sigma_2^2},\ df_{числ} = N_1 - 1;\ df_{знам} = N_2 - 1\ , \qquad (11.1)$$

где σ_1^2 — большая дисперсия, а σ_2^2 — меньшая дисперсия. Так как заранее не известно, какая дисперсия больше, то для определения p-уровня применяется *Таблица критических значений для ненаправленных альтернатив*. Если $F_э \geqslant F_{кр}$ для соответствующего числа степеней свободы, то $p \leqslant 0{,}05$ и статистическую гипотезу о равенстве дисперсий можно отклонить (для $\alpha = 0{,}05$).

Метод может применяться для проверки предположения о равенстве (гомогенности) дисперсий перед проверкой достоверности различия средних по критерию t-Стьюдента для независимых выборок разной численности. Однако содержательная интерпретация статистически достоверного различия дисперсий может иметь и самостоятельную ценность.

ПРИМЕР 11.1[1]

Детям давались обычные арифметические задания, после чего одной случайно выбранной половине учащихся сообщали, что они не выдержали испытания, а остальным — обратное. Затем у каждого ребенка спрашивали, сколько секунд ему потребовалось бы для решения аналогичной задачи. Экспериментатор вычислял разность между называемым ребенком временем и результатом выполненного задания (в сек.). Ожидалось, что сообщение о неудаче вызовет некоторую неадекватность самооценки ребенка. Проверяемая гипотеза (на уровне $\alpha = 0{,}05$) состояла в

[1] *Гласс Дж., Стэнли Дж.* Статистические методы в педагогике и психологии. М., 1976. С. 277.

том, что дисперсия совокупности самооценок не зависит от сообщений об удаче или неудаче (H_0: $\bar{\sigma}_1^2 = \bar{\sigma}_2^2$).

Были получены следующие данные:

Группа 1: сообщение о неудаче	Группа 2: сообщение об успехе
$N_1 = 12$ $\sigma_1^2 = 90{,}45$	$N_2 = 12$ $\sigma_2^2 = 8{,}16$

Ш а г 1. Вычислим эмпирическое значение критерия и числа степеней свободы по формулам 11.1:

$$F_э = \frac{90{,}45}{8{,}16} = 11{,}08,\ df_{числ} = 11;\ df_{знам} = 11.$$

Ш а г 2. По таблице критических значений критерия *F*-Фишера для *ненаправленных* альтернатив (приложение 8) находим критическое значение для $df_{числ} = 11$; $df_{знам} = 11$. Однако критическое значение есть только для $df_{числ} = 10$ и $df_{числ} = 12$. Большее число степеней свободы брать нельзя, поэтому берем критическое значение для $df_{числ} = 10$: для $p = 0{,}05$ $F_{кр} = 3{,}526$; для $p = 0{,}01$ $F_{кр} = 5{,}418$.

Ш а г 3. Принятие статистического решения и содержательный вывод. Поскольку эмпирическое значение превышает критическое значение для $p = 0{,}01$ (и тем более — для $p = 0{,}05$), то в данном случае $p < 0{,}01$ и принимается альтернативная гипотеза: дисперсия в группе 1 превышает дисперсию в группе 2 ($p < 0{,}01$). Следовательно, после сообщения о неудаче неадекватность самооценки выше, чем после сообщения об удаче.

КРИТЕРИЙ *t*-СТЬЮДЕНТА ДЛЯ ОДНОЙ ВЫБОРКИ

Метод позволяет проверить гипотезу о том, что среднее значение изучаемого признака M_x отличается от некоторого известного значения *A*. Проверяемая статистическая гипотеза: H_0: $\bar{M}_x = A$. При ее отклонении принимается альтернативная гипотеза о том, что M_x меньше (больше) *A*.

Исходное предположение: распределение признака в выборке приблизительно соответствует нормальному виду.

Структура исходных данных: значения изучаемого признака определены для каждого члена выборки, которая репрезентативна изучаемой генеральной совокупности.

Альтернатива методу: нет.

Формула для эмпирического значения критерия *t*-Стьюдента:

$$t_э = \frac{|M - A|}{\sigma/\sqrt{N}},\ df = N - 1. \quad (11.2)$$

ПРИМЕР 11.2

Предположим, исследовалось влияние условий воспитания в детском доме на интеллектуальное развитие детей. При использовании стандартного теста интеллекта для случайной выборки воспитанников детдома были получены следующие резуль-

таты: $M = 106$; $\sigma = 15$; $N = 36$. Исследователя интересовало, превышает ли интеллект воспитанников детдома нормативный показатель $A = 100$. Для принятия статистического решения был определен уровень $\alpha = 0,05$.

Шаг 1. Вычисляем по формуле 11.2 эмпирическое значение критерия и число степеней свободы: $t_э = 2,4$; $df = 35$.

Шаг 2. Определяем по таблице критических значений критерия t-Стьюдента (приложение 2) p-уровень значимости. Для $df = 35$ эмпирическое значение находится между критическими для $p = 0,05$ и $p = 0,01$. Следовательно, $p < 0,05$.

Шаг 3. Принимаем статистическое решение и формулируем вывод. Статистическая гипотеза о равенстве среднего значения заданной величине отклоняется. Интеллект воспитанников детдома ($M = 106$; $\sigma = 15$; $N = 36$) статистически достоверно превышает нормативный показатель интеллекта $A = 100$ ($p < 0,05$).

КРИТЕРИЙ t-СТЬЮДЕНТА ДЛЯ НЕЗАВИСИМЫХ ВЫБОРОК

Метод позволяет проверить гипотезу о том, что средние значения двух генеральных совокупностей, из которых извлечены сравниваемые *независимые* выборки, отличаются друг от друга. Допущение независимости предполагает, что представители двух выборок *не составляют пары коррелирующих значений признака*. Это предположение нарушилось бы, если, например, 1-я выборка состояла из мужей, а 2-я — из их жен, и два ряда значений измеренного признака могли бы коррелировать.

Проверяемая статистическая гипотеза H_0: $\bar{M}_1 = \bar{M}_2$. При ее отклонении принимается альтернативная гипотеза о том, что M_1 больше (меньше) M_2.

Исходные предположения для статистической проверки:

- одна выборка извлекается из одной генеральной совокупности, а другая выборка, *независимая* от первой, извлекается из другой генеральной совокупности;
- распределение изучаемого признака и в той, и в другой выборке приблизительно соответствует нормальному;
- дисперсии признака в двух выборках примерно одинаковы (гомогенны).

Структура исходных данных: изучаемый признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из двух сравниваемых независимых выборок.

Ограничения: распределения признака и в той, и в другой выборке существенно не отличаются от нормального; *в случае разной численности* сравниваемых выборок их дисперсии статистически достоверно не различаются (проверяется по критерию F-Фишера — при вычислениях «вручную», по критерию Ливена — при вычислениях на компьютере).

Альтернатива методу: непараметрический критерий *U*-Манна-Уитни — если распределение признака хотя бы в одной выборке существенно отличается от нормального и (или) дисперсии различаются статистически достоверно.

Формулы для эмпирического значения критерия *t*-Стьюдента:

$$t_э = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \qquad (11.3)$$

или

$$t_э = \frac{|M_1 - M_2|}{\sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2}\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}} \qquad (11.4)$$

$$df = N_1 + N_2 - 2 .$$

Формула (11.3) применяется для приближенных расчетов, для близких по численности выборок, а формула (11.4) — для точных расчетов, когда выборки заметно различаются по численности.

ПРИМЕР 11.3

Предположим, изучалось различие в интеллекте студентов 1-го и 5-го курсов. Для этого случайным образом были отобраны 30 студентов 1 курса и 28 студентов 5 \курса, у которых интеллект определялся по одной и той же методике. Были получены следующие результаты:

Группа 1: 1-й курс	Группа 2: 5-й курс
$N_1 = 30$	$N_2 = 28$
$M_1 = 103$	$M_2 = 109$
$\sigma_1 = 10$	$\sigma_2^2 = 12$

Гипотеза о различии интеллекта проверялась на уровне $\alpha = 0{,}05$.

Шаг 1. Вычисляем эмпирическое значение критерия *t*-Стьюдента по формуле 11.3: $t_э = 2{,}06$ (по формуле 11.4: $t_э = 2{,}17$); $df = 56$.

Шаг 2. Определяем по таблице критических значений критерия *t*-Стьюдента (приложение 2) *p*-уровень значимости. Для $df = 56$ эмпирическое значение находится между критическими для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$.

Шаг 3. Принимаем статистическое решение и формулируем вывод. Статистическая гипотеза о равенстве средних значений отклоняется. Вывод: интеллект студентов 5 курса статистически достоверно выше, чем у студентов 1 курса ($p < 0{,}05$).

КРИТЕРИЙ *t*-СТЬЮДЕНТА ДЛЯ ЗАВИСИМЫХ ВЫБОРОК

Метод позволяет проверить гипотезу о том, что средние значения двух генеральных совокупностей, из которых извлечены сравниваемые *зависимые* выборки, отличаются друг от друга. Допущение зависимости чаще всего значит, что признак измерен на одной и той же выборке дважды, например, до воздействия и после него. В общем же случае каждому представителю одной выборки поставлен в соответствие представитель из другой выборки (они попарно объединены) так, что два ряда данных положительно коррелируют друг с другом. Более слабые виды зависимости выборок: выборка 1 — мужья, выборка 2 — их жены; выборка 1 — годовалые дети, выборка 2 составлена из близнецов детей выборки 1, и т. д.

Проверяемая статистическая гипотеза, как и в предыдущем случае, H_0: $\bar{M}_1 = \bar{M}_2$. При ее отклонении принимается альтернативная гипотеза о том, что $\bar{M}_1$ больше (меньше) $\bar{M}_2$.

Исходные предположения для статистической проверки:

- каждому представителю одной выборки (из одной генеральной совокупности) поставлен в соответствие представитель другой выборки (из другой генеральной совокупности);
- данные двух выборок положительно коррелируют;
- распределение изучаемого признака и в той и другой выборке соответствует нормальному закону.

Структура исходных данных: имеется по два значения изучаемого признака для каждого объекта (для каждой пары).

ПРИМЕР

При сравнении значений признака X до воздействия (X_1) и после воздействия (X_2) на выборку численностью N:

№	X_1	X_2
1	8	10
2	8	9
3	3	4
4	5	5
...	...	...
N	6	7

Ограничения: распределения признака и в той, и в другой выборке существенно не отличаются от нормального; данные двух измерений, соответствующих той и другой выборке, положительно коррелируют.

Альтернативы: критерий T-Вилкоксона, если распределение хотя бы для одной выборки существенно отличается от нормального; критерий t-Стью-

дента для независимых выборок — если данные для двух выборок не коррелируют положительно.

Формула для эмпирического значения критерия t-Стьюдента отражает тот факт, что единицей анализа различий является *разность* (*сдвиг*) значений признака для каждой пары наблюдений. Соответственно, для каждой из N пар значений признака сначала вычисляется разность $d_i = x_{1i} - x_{2i}$.

$$t_э = \frac{|M_d|}{\sigma_d/\sqrt{N}},\ df = N - 1, \qquad (11.5)$$

где M_d — средняя разность значений; σ_d — стандартное отклонение разностей.

ПРИМЕР 11.4

Предположим, в ходе проверки эффективности тренинга каждому из 8 членов группы задавался вопрос «Насколько часто твое мнение совпадает с мнением группы?» — дважды, до и после тренинга. Для ответов использовалась 10-балльная шкала: 1 — никогда, ..., 5 — в половине случаев, ..., 10 — всегда. Проверялась гипотеза о том, что в результате тренинга самооценка конформизма участников возрастет ($\alpha = 0{,}05$).

Составим таблицу для промежуточных вычислений:

№	X_1	X_2	$d_i = X_1 - X_2$	$d_i - M_d$	$(d_i - M_d)2$
1	3	4	–1	–0,25	0,0625
2	6	6	0	0,75	0,5625
3	5	6	–1	–0,25	0,0625
4	2	4	–2	–1,25	1,5625
5	7	6	1	1,75	3,0625
6	3	4	–1	–0,25	0,0625
7	4	5	–1	–0,25	0,0625
8	5	6	–1	–0,25	0,0625
Сумма:	35	41	–6	0	5,5

Шаг 1. Вычисляем эмпирическое значение критерия по формуле 11.5: средняя разность $M_d = -0{,}75$; стандартное отклонение $\sigma_d = 0{,}886$; $t_э = 2{,}39$; $df = 7$.

Шаг 2. Определяем по таблице критических значений критерия t-Стьюдента (приложение 2) p-уровень значимости. Для $df = 7$ эмпирическое значение находится между критическими для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$.

Шаг 3. Принимаем статистическое решение и формулируем вывод. Статистическая гипотеза о равенстве средних значений отклоняется. Вывод: показатель самооценки конформизма участников после тренинга увеличился статистически достоверно ($p < 0{,}05$).

Замечание. В отношении зависимых выборок вполне допустимо применение критерия t-Стьюдента для независимых выборок (но не наоборот!). Это целесообразно, если корреляция между двумя измерениями отрицательная. Если же корреляция положительная, то такая замена приведет к недооценке достоверности различий.

В примере 11.4 корреляция между X_1 и X_2 $r = 0,9$. Если в отношении данных применить формулу 11.3, то эмпирическое значение критерия составит $t_э = 1,085$. Для $df = 14$ это значение значительно меньше критического для $p = 0,1$. Следовательно, статистическая гипотеза о равенстве средних значений не отклоняется.

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Критерий *t*-Стьюдента для одной выборки.

А) Выбираем **Analyze > Compare means > One Sample T-Test...**

Б) В открывшемся окне диалога выделяем и переносим интересующие переменные из левого окна в правое окно при помощи кнопки ▷ (в данном случае — переменные var7 и var8). Устанавливаем величину, с которой собираемся сравнивать средние значения: **Test Value** — вводим значение (в данном примере 10). Нажимаем OK.

В) Получаем результаты в виде двух таблиц:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
VAR7	46	10.9130	2.88156	.42486
VAR8	46	9.6957	2.50217	.36893

One-Sample Test

	Test Value = 10			
	t	df	Sig. (2-tailed)	Mean Difference
VAR7	2.149	45	.037	.9130
VAR8	-.825	45	.414	-.3043

В первой таблице содержатся первичные статистики, в частности, средние значения (Means), стандартные отклонения (Std. Deviation). Во второй — результаты проверки гипотез: значения *t*-Стьюдента (*t*), числа степеней свободы (*df*), уровень значимости (Sig.), разность среднего значения и заданной величины (Mean Difference).

Критерий *t*-Стьюдента и сравнение двух дисперсий для независимых выборок.

А) Выбираем **Analyze > Compare means > Independent Samples T-Test...**

Б) В открывшемся окне диалога выделяем и переносим при помощи кнопки ▷ из левого окна интересующие переменные в правое верхнее окно (**Test Variable(s)**) (в данном случае — переменную var1); группирующую переменную, которая делит выборку на подгруппы (**Grouping Variable**) (в данном случае — переменную var2). Нажимаем кнопку **Define Groups...** и задаем номера градаций группирующей переменной, которые мы хотим сравнить (в данном случае 0 и 1). Нажимаем **Continue**. Нажимаем OK.

В) Получаем результаты в виде двух таблиц.

Group Statistics

	VAR2	N	Mean	Std. Deviation	Std. Error Mean
VAR1	.00	11	9.1818	2.99393	.90270
	1.00	9	13.3333	4.33013	1.44338

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means		
		F	Sig.	t	df	Sig. (2-tailed)
VAR1	Equal variances assumed	1.344	.262	-2.531	18	.021
	Equal variances not assumed			-2.439	13.794	.029

В первой таблице содержатся первичные статистики: каждой выборке (градации 0 и 1 переменной var2) соответствует своя строка. Во второй таблице — результаты проверки гипотез: проверка равенства дисперсий (**Levene's Test...**) — значение критерия (**F**) и уровень значимости (**Sig.**); проверка различий средних (**t-test...**) — значение критерия (***t***), число степеней свободы (***df***), уровень значимости (**Sig.**). Результаты проверки по критерию *t*-Стьюдента принимаются во внимание, если дисперсии по критерию Ливена статистически значимо не различаются ($p > 0,05$).

Примечание. Программа предлагает вариант проверки по критерию *t*-Стьюдента с поправкой на неоднородность дисперсии (**Equal variances not assumed**). Правомерность такой поправки не очевидна. Поэтому, в случае получения достоверного различия дисперсий по критерию Ливена, рекомендуем воспользоваться непараметрической проверкой различий для независимых выборок (по критерию *U*-Манна-Уитни).

Критерий *t*-Стьюдента для зависимых выборок.

А) Выбираем **Analyze > Compare means > Paired-Samples T-Test...**

Б) В открывшемся окне диалога выделяем две переменные (соответствующие двум зависимым выборкам — измерениям одного и того же признака) и переносим пару при помощи кнопки ▷ из левого окна в правое окно (**Paired Variables**). Пар может быть несколько (в данном случае — это пара переменных var2 и var3). Нажимаем ОК.

В) Получаем результаты в виде трех таблиц:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	VAR2	11.9000	20	2.46875	.55203
	VAR3	9.6000	20	3.15228	.70487

Paired Samples Correlations

	N	Correlation	Sig.
Pair 1 VAR2 & VAR3	20	.482	.032

Paired Samples Test

	Paired Differences					
	Mean	Std. Deviation	Std. Error Mean	t	df	Sig. (2-tailed)
Pair 1 VAR2 — VAR3	2.30	2.922	.65333	3.520	19	.002

Первая таблица содержит первичные статистики: каждой выборке соответствует своя строка. Во второй таблице — корреляция Пирсона для пары переменных, соответствующих двум зависимым выборкам. Третья таблица содержит результаты проверки гипотезы по критерию *t*-Стьюдента: среднюю разность (Mean), стандартное отклонение разности (**Std. Deviation**), значение критерия (***t***), число степеней свободы (***df***), уровень значимости (**Sig.**).

Глава 12

НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ ВЫБОРОК

ОБЩИЕ ЗАМЕЧАНИЯ

К методам сравнения выборок, в соответствии с принятой классификацией[1], мы относим способы проверки статистических гипотез о различии выборок по уровню выраженности признака, измеренного в количественной шкале. Непараметрические методы сравнения выборок, рассматриваемые в этой главе, являются аналогами параметрических методов сравнения средних значений. И почти каждый параметрический метод сравнения средних может быть при необходимости заменен своим непараметрическим аналогом либо сочетанием непараметрических методов[2].

Непараметрические методы заметно проще в вычислительном отношении, чем их параметрические аналоги. До недавнего прошлого простота вычислений имела существенное значение при обработке данных «вручную». Но, во-первых, данные очень часто включают одинаковые значения, усложняющие расчеты, во-вторых, компьютерная обработка снимает проблему сложности вычислений. Поэтому при выборе между параметрическими и непараметрическими методами следует исходить из свойств самих данных.

Непараметрические аналоги параметрических методов сравнения выборок применяются в случаях, когда не выполняются основные предположения, лежащие в основе параметрических методов сравнения средних значений.

При решении вопроса о выборе параметрического или непараметрического метода сравнения необходимо иметь в виду, что параметрические методы обладают заведомо большей чувствительностью, чем их непараметрические аналоги. Поэтому исходной ситуацией является выбор параметрического метода. И решение о применении непараметрического метода становится оп-

---

[1] Сравнение выборок по качественно определенному (номинативному) признаку мы относим к задачам анализа классификаций и таблиц сопряженности.

[2] Данная глава содержит описание не всех, но самых популярных непараметрических методов. Более полный их арсенал, вместе с подробными рекомендациями по применению и с богатым иллюстративным материалом, можно найти в кн.: *Сидоренко Е. В.* Методы математической обработки в психологии. СПб., 1996.

равданным, если не выполняются исходные предположения, лежащие в основе применения параметрического метода.

Условия, когда применение непараметрических методов является оправданным:

- есть основания считать, что распределение значений признака в генеральной совокупности не соответствует нормальному закону;
- есть сомнения в нормальности распределения признака в генеральной совокупности, но выборка слишком мала, чтобы по выборочному распределению судить о распределении в генеральной совокупности;
- не выполняется требование гомогенности дисперсии при сравнении средних значений для независимых выборок.

На практике *преимущество непараметрических методов* наиболее заметно, когда в данных имеются выбросы (экстремально большие или малые значения).

Если размер выборки очень велик (больше 100), то непараметрические методы сравнения использовать нецелесообразно, даже если не выполняются некоторые исходные предположения применения параметрических методов. С другой стороны, если объемы сравниваемых выборок очень малы (10 и меньше), то результаты применения непараметрических методов можно рассматривать лишь как предварительные.

Структура исходных данных и интерпретация результатов применения для параметрических методов и их непараметрических аналогов являются идентичными.

При сравнении выборок с использованием непараметрических критериев, как и в случае параметрических критериев, обычно проверяются *ненаправленные статистические гипотезы.* Основная (нулевая) статистическая гипотеза при этом содержит утверждение об идентичности генеральных совокупностей (из которых извлечены выборки) по уровню выраженности изучаемого признака. Соответственно, при ее отклонении допустимо принятие двусторонней альтернативы о конкретном направлении различий в соответствии с выборочными данными. Для принятия статистического решения в таких случаях применяются двусторонние критерии и, соответственно, критические значения для проверки ненаправленных альтернатив.

Перед знакомством с непараметрическими методами сравнения читателю необходимо ознакомиться с порядком и условиями применения их параметрических аналогов.

При выборе того или иного непараметрического метода сравнения выборок можно руководствоваться таблицей классификации методов сравнения (см. рис. 8.2).

СРАВНЕНИЕ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК

Самым популярным и наиболее чувствительным (мощным) аналогом критерия *t*-Стьюдента для независимых выборок является *критерий U-Манна-Уитни (Mann-Whitney U).* Непараметрическим его аналогом является *крите-*

рий серий (см. главу 8), который еще проще в вычислительном отношении, но обладает заметно меньшей чувствительностью, чем критерий *U*.

Эмпирическое значение критерия *U*-Манна-Уитни показывает, насколько совпадают (пересекаются) два ряда значений измеренного признака. Чем меньше совпадение, тем больше различаются эти два ряда. Основная идея критерия *U* основана на представлении всех значений двух выборок в виде одной общей последовательности упорядоченных (ранжированных) значений. Основной (нулевой) статистической гипотезе будет соответствовать ситуация, когда значения одной выборки будут равномерно распределены среди значений другой выборки, то есть когда два ряда значений пересекаются в наибольшей возможной степени. Напротив, отклонению этой гипотезы будет соответствовать ситуация, когда значения одной из выборок будут преобладать на одном из концов объединенного ряда — пересечение двух рядов тогда будет минимальным.

ПРИМЕР 12.1

Обозначим значения переменной для одной выборки *X*, а для другой выборки — *Y* и упорядочим значения обеих выборок по возрастанию.

Значения	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19
Выборка	*X*	*X*	*Y*	*X*	*X*	*X*	*Y*	*X*	*X*	*Y*	*X*	*Y*	*Y*	*Y*	*Y*	*Y*

Значения одной выборки распределены явно не равномерно среди значений другой выборки: значения выборки *Y* преобладают на правом конце объединенного ряда. Однако критерий серий не позволяет обнаружить статистически значимые различия: всего серий в данном случае $W = 8$ и при $m = n = 8$ эта величина не выходит за пределы критических значений для $\alpha = 0{,}05$ (приложение 5).

Формально, критерий *U* — это общее число тех случаев, в которых значения одной группы превосходят значения другой группы, при попарном сравнении значений первой и второй групп. Соответственно, вычисляются два значения критерия: U_x и U_y.

Для вычислений «вручную» используются следующие формулы:

$$U_x = mn - R_x + \frac{n(n+1)}{2},$$
$$U_y = mn - R_y + \frac{m(m+1)}{2}, \qquad (12.1)$$
$$U_x + U_y = mn,$$

где n — объем выборки *X*; m — объем выборки *Y*, R_x и R_y — суммы рангов для *X* и *Y* в объединенном ряду. В качестве эмпирического значения критерия берется *наименьшее* из U_x и U_y. Чем больше различия, тем меньше эмпирическое значение *U*.

Поскольку критерий *U* отражает степень совпадения (перекрещивания) двух рядов значений, то *значение p-уровня тем меньше, чем меньше значение U*. При расчетах «вручную» используют таблицы критических значений критерия *U*-Манна-Уитни (приложение 9).

ПРИМЕР 12.1 (продолжение)

Проверим гипотезу о различии выборок X (численностью $m = 8$) и Y (численностью $n = 8$) на уровне $\alpha = 0{,}05$:

1	Значения	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19
2	Выборка	X	X	Y	X	X	X	Y	X	X	Y	X	Y	Y	Y	Y	Y
3	Ранги	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4	Ранги X	1	2		4	5	6		8	9		11					
5	Ранги Y			3				7			10		12	13	14	15	16

Шаг 1. Значения двух выборок объединяются в один ряд, упорядоченный в порядке возрастания или убывания. Обозначается принадлежность каждого значения к той и другой выборке (строки 1 и 2).

Шаг 2. Значения выборок ранжируются, и выписываются отдельно ранги для одной и другой выборки (строки 3–5).

Шаг 3. Вычисляются суммы рангов по $X(R_x)$ и по $Y(R_y)$: $R_x = 46$; $R_y = 90$.

Шаг 4. Вычисляются U_x и U_y по формуле 12.1:

$$U_x = 8 \cdot 8 - 46 + \frac{8(8+1)}{2} = 54, \quad U_y = 8 \cdot 8 - 90 + \frac{8(8+1)}{2} = 10, \quad U_x + U_y = 64 = mn.$$

Шаг 5. Определяется p-уровень значимости: наименьшее из U сравнивается с табличным (приложение 9) для соответствующих объемов выборки $m = 8$ и $n = 8$. Значение $p \leqslant 0{,}05$ ($0{,}01$), если вычисленное $U_{\text{эмп.}} \leqslant U_{\text{табл.}}$ В нашем случае наименьшим является $U_y = 10$, которое и принимается за эмпирическое значение критерия. Оно меньше критического для $p = 0{,}05$ ($U = 13$), но больше критического для $p = 0{,}01$ ($U = 7$). Следовательно, $p < 0{,}05$.

Шаг 6. Принимается статистическое решение и формулируется содержательный вывод. На уровне $\alpha = 0{,}05$ принимается статистическая гипотеза о различии X и Y по уровню выраженности признака. Уровень Y статистически достоверно выше уровня X ($p < 0{,}05$).

Замечание. Связи в рангах для вычислений «вручную» не предусмотрены. Хотя они и незначительно влияют на результат, но если доля одинаковых рангов по одной из переменных велика, то предлагаемый алгоритм неприменим, пользуйтесь компьютерной программой (SPSS, Statistica).

Обработка на компьютере: критерий U-Манна-Уитни

Для обработки использованы данные примера 12.1. В таблице исходных данных (**Data Editor**) для каждого из 16 объектов определены значения двух переменных: `var1` – значения количественного признака, `var2` – бинарная группирующая переменная, обозначающая принадлежность каждого объекта к одной из двух групп.

А) Выбираем **Analyze > Nonparametric Tests > 2-Independent Samples...** (Две независимые выборки).

Б) В открывшемся окне диалога выделяем и переносим при помощи кнопки ▷ из левого окна интересующие переменные (в данном случае var1) в правое верхнее окно (**Test Variable(s)**); группирующую переменную (в данном случае var2), которая делит выборку на подгруппы (**Grouping Variable**). Нажимаем кнопку **Define Groups...** и задаем номера градаций группирующей переменной, которые мы хотим сравнить (1 и 2). Нажимаем **Continue**. Нажимаем OK.

В) Получаем результаты в виде двух таблиц:

Ranks

	VAR2	N	Mean Rank	Sum of Ranks
VAR1	1.00	8	5.75	46.00
	2.00	8	11.25	90.00
	Total	16		

Test Statistics(b)

	VAR1
Mann-Whitney U	10.000
Wilcoxon W	46.000
Z	-2.310
Asymp. Sig. (2-tailed)	.021
Exact Sig. [2*(1-tailed Sig.)]	.021(a)

a Not corrected for ties.
b Grouping Variable: VAR2

В первой таблице содержатся ранговые статистики: средние ранги для групп (**Mean Rank**) и суммы рангов (**Sum of Ranks**). Во второй таблице содержатся результаты проверки гипотезы: эмпирическое значение *U*-критерия (**Mann-Whitney U**) и *p*-уровень значимости (**Asymp. Sig. (2-tailed)**).

СРАВНЕНИЕ ДВУХ ЗАВИСИМЫХ ВЫБОРОК

Самым чувствительным (мощным) аналогом критерия *t*-Стьюдента для зависимых выборок является *критерий Т-Вилкоксона (Wilcoxon signed-rank test)*. Непараметрическим его аналогом является *критерий знаков*, который еще проще в вычислительном отношении, но обладает меньшей чувствительностью, чем критерий *Т*-Вилкоксона. Критерий *Т* основан на упорядочивании величин *разностей* (сдвигов) значений признака в каждой паре его измерений (критерий знаков основан на учете только знака этой разности). Соответственно, критерий *Т*, будучи менее чувствительным аналогом *t*-Стьюдента, более чувствителен по сравнению с другими непараметрическими критериями для повторных измерений (зависимых выборок).

Т-Вилкоксона основан на ранжировании абсолютных разностей пар значений зависимых выборок. Далее подсчитывается сумма рангов для положительных разностей и сумма рангов для отрицательных разностей. Идея критерия *Т* заключается в подсчете вероятности получения минимальной из этих разностей при условии, что распределение положительных или отрицательных разностей равновероятно и равно 1/2.

Для расчетов «вручную» не требуется особых формул: достаточно подсчитать суммы рангов для положительных и отрицательных разностей. Затем меньшая из сумм принимается в качестве эмпирического значения критерия, значение которого сравнивается с табличным значением (приложение 10), рассчитанным для условия равной вероятности положительных и отрицательных разностей для данного объема выборки. Конечно, чем больше различия, тем меньше эмпирическое значение *Т*, тем менее вероятно получение такого значения при условии равной вероятности встречаемости положительных и отрицательных разностей, следовательно, тем меньше значение *р*-уровня.

ПРИМЕР 12.2

Проверим гипотезу о различии значений показателя, измеренного дважды на одной и той же выборке («Условие 1» и «Условие 2»), на уровне $\alpha = 0{,}05$:

1	№ объекта:	1	2	3	4	5	6	7	8	9	10	11	12
2	Условие 1:	6	11	12	8	5	10	7	6	3	9	4	5
3	Условие 2:	14	5	8	10	14	7	12	13	11	10	15	16
4	Разность d_i:	–8	6	4	–2	–9	3	–5	–7	–8	–1	–11	–11
5	Ранги $\|d_i\|$:	8,5	6	4	2	10	3	5	7	8,5	1	11,5	11,5
6	Ранги d_i (+):		6	4			3						
7	Ранги d (–):	8,5			2	10		5	7	8,5	1	11,5	11,5

Шаг 1. Подсчитать разности значений для каждого объекта выборки (строка 4).

Шаг 2. Ранжировать абсолютные значения разностей (строка 5).

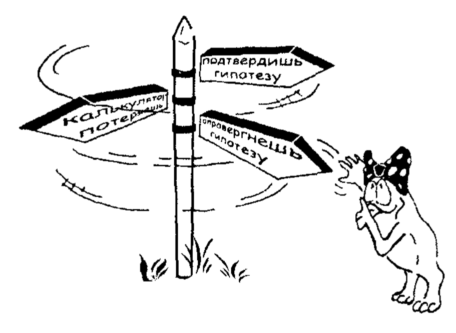
Шаг 3. Выписать ранги положительных и отрицательных значений разностей (строки 6 и 7).

Шаг 4. Подсчитать суммы рангов отдельно для положительных и отрицательных разностей: $T_1 = 13$; $T_2 = 65$. За эмпирическое значение критерия $T_{эмп.}$ принимается меньшая сумма: $T_{эмп} = 13$.

Шаг 5. Определяется *р*-уровень значимости: $T_{эмп}$ сравнивается с табличным (приложение 10) для соответствующего объема выборки. Значение $p \leqslant 0{,}05$ (0,01), если вычисленное $T_{эмп} \leqslant T_{табл}$ В нашем случае эмпирическое значение равно критическому значению для $p = 0{,}05$. Следовательно, $p = 0{,}05$.

Шаг 6. Принимается статистическое решение и формулируется содержательный вывод. На уровне $\alpha = 0{,}05$ принимается статистическая гипотеза о различии двух условий по уровню выраженности изучаемого признака. Уровень выраженности признака для условия 2 статистически значимо выше, чем для условия 1 ($p = 0{,}05$).

Замечание. Связи в рангах абсолютных значений разностей для вычислений «вручную» не предусмотрены. Хотя их влияние и не очень суще-



ственно, но если доля одинаковых рангов велика и превышает, скажем, 50%, то предлагаемый алгоритм неприменим, пользуйтесь компьютерной программой (SPSS, Statistica) или *G*-критерием знаков.

Критерий знаков G (Sign test) — менее чувствительная к сдвигам альтернатива критерия *T*-Вилкоксона. Для того чтобы им воспользоваться, достаточно подсчитать количество отрицательных и положительных сдвигов.

ПРИМЕР

Проверим гипотезу о различии в отношении данных примера 12.2 с использованием критерия знаков (на уровне $\alpha = 0{,}05$).

Шаг 1. Подсчитать количество положительных и отрицательных разностей значений (по строке 4). Сдвиг в значениях, соответствующий наибольшему числу из этих разностей, принимается за типичный сдвиг. Количество типичных сдвигов обозначается N, а количество нетипичных сдвигов принимается в качестве эмпирического значения критерия $G_{эмп}$ В нашем случае количество типичных сдвигов $N = 9$, а количество нетипичных сдвигов $G_{эмп} = 3$.

Шаг 2. Определяется p-уровень значимости: $G_{эмп}$ (количество нетипичных сдвигов) сравнивается с табличным критическим (приложение 11) для соответствующего N (количества типичных сдвигов). Чем меньше $G_{эмп}$, тем меньше значение p-уровня. Значение $p \leqslant 0{,}05$ (0,01), если вычисленное $G_{эмп} \leqslant G_{табл}$ В нашем случае для $N = 9$ табличное значение для $p = 0{,}05$ равно 1, и $G_{эмп}$ его превышает. Следовательно, $p > 0{,}05$.

Шаг 3. Принимается статистическое решение и формулируется содержательный вывод. На уровне $\alpha = 0{,}05$ принимается нулевая статистическая гипотеза об отсутствии различий. Между условиями 1 и 2 не обнаружены статистически достоверные различия в уровне выраженности изучаемого признака ($p > 0{,}05$).

Обработка на компьютере: критерий *T*-Вилкоксона

Для обработки использованы данные примера 12.2. Исходные данные для обработки введены в таблицу (**Data Editor**) в виде двух переменных: `var1` – «Условие 1»; `var2` – «Условие 2».

А) Выбираем **Analyze > Nonparametric Tests > 2-Related Samples...** (Две зависимые выборки).

Б) В открывшемся окне диалога выделяем две переменные (соответствующие двум измерениям одного и того же признака) и переносим пару при помощи кнопки ▷ из левого окна в правое окно (**Paired Variables**). Пар может быть несколько. Нажимаем ОК.

В) Получаем результаты в виде двух таблиц:

Ranks

		N	Mean Rank	Sum of Ranks
VAR1 - VAR2	Negative Ranks	3(a)	4.33	13.00
	Positive Ranks	9(b)	7.22	65.00
	Ties	0(c)		
	Total	12		

a VAR00004 < VAR00003
b VAR00004 > VAR00003
c VAR00004 = VAR00003

Test Statistics(b)

	VAR1 - VAR2
Z	-2.041(a)
Asymp. Sig. (2-tailed)	.041

a Based on negative ranks.
b Wilcoxon Signed Ranks Test

В первой таблице содержатся ранговые статистики: средние ранги (**Mean Rank**) и суммы рангов (**Sum of Ranks**) для отрицательных (**Negative Ranks**) и положительных (**Positive Ranks**) сдвигов, а также количество одинаковых рангов (**Ties**). Во второй таблице содержатся результаты проверки гипотезы: эмпирическое значение z-критерия (**Z**) и p-уровень значимости (**Asymp. Sig. (2-tailed)**).

СРАВНЕНИЕ БОЛЕЕ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК

Критерий Н Краскала-Уоллеса (Kruskal-Wallis H) является непараметрическим аналогом однофакторного дисперсионного анализа (ANOVA) для независимых выборок, поэтому другое его название — *Однофакторный дисперсионный анализ Краскала-Уоллеса (Kruskal-Wallis one-way analysis of variance).* Он позволяет проверять гипотезы о различии более двух выборок по уровню выраженности изучаемого признака.

Н-Краскала-Уоллеса по идее сходен с критерием *U*-Манна-Уитни. Как и последний, он оценивает степень пересечения (совпадения) нескольких рядов значений измеренного признака. Чем меньше совпадений, тем больше различаются ряды, соответствующие сравниваемым выборкам. Основная идея критерия *Н*-Краскала-Уоллеса основана на представлении всех значений сравниваемых выборок в виде одной общей последовательности упорядоченных (ранжированных) значений, с последующим вычислением среднего ранга для

каждой из выборок. Если выполняется статистическая гипотеза об отсутствии различий, то можно ожидать, что все средние ранги примерно равны и близки к общему среднему рангу.

Эмпирическое значение критерия H-Краскала-Уоллеса вычисляется после ранжирования всех значений сравниваемых выборок по формуле:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1), \quad (12.2)$$

где N — суммарная численность всех выборок; k — количество сравниваемых выборок; R_i — сумма рангов для выборки i; n_i — численность выборки i. Чем сильнее различаются выборки, тем больше вычисленное значение H и тем меньше p-уровень значимости.

При расчетах «вручную» для определения p-уровня пользуются таблицами критических значений. Если объем каждой выборки больше 5 и количество выборок больше трех, то эмпирическое значение критерия сравнивается с χ^2 (приложение 4) для $df = k - 1$ (k — число выборок). Если сравниваются 3 выборки и объем каждой выборки меньше 5, то пользуются таблицей критических значений H-Краскала-Уоллеса (приложение 12).

При отклонении нулевой статистической гипотезы об отсутствии различий принимается альтернативная гипотеза о статистически достоверных различиях выборок по изучаемому признаку — без конкретизации направления различий. *Для утверждений о том, что уровень выраженности признака в какой-то из сравниваемых выборок выше или ниже, необходимо парное соотнесение выборок по критерию U-Манна-Уитни.*

ПРИМЕР 12.3

Проверим гипотезу о различии выборок 1, 2 и 3 на уровне $\alpha = 0{,}05$:

1	Значения	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	19
2	Выборка	1	1	2	1	1	1	2	1	1	2	1	3	2	3	3	2
3	Ранги	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
4	Ранги 1	1	2		4	5	6		8	9		11					
5	Ранги 2			3				7			10			13			16
6	Ранги 3												12		14	15	

Шаг 1. Значения выборок объединяются в один ряд, упорядоченный в порядке возрастания или убывания. Обозначается принадлежность каждого значения к той или иной выборке (строки 1 и 2).

Шаг 2. Значения выборок ранжируются и выписываются отдельно ранги для каждой выборки (строки 3–6).

Шаг 3. Вычисляются суммы рангов для каждой выборки и проверяется правильность расчетов. $R_1 = 46$; $R_2 = 49$; $R_3 = 41$. Общая сумма рангов должна быть равна $N(N+1)/2 = 16 \times 17/2 = 136$. Равенство соблюдено.

Шаг 4. Вычисляется H по формуле 12.2:

$$H = \frac{12}{16(16+1)}\left(\frac{46^2}{8} + \frac{49^2}{5} + \frac{41^2}{3}\right) - 3(16+1) = 6{,}575\,.$$

Шаг 5. Определяется p-уровень значимости. Хотя сравниваются 3 выборки, но объем одной из них больше 5, поэтому вычисленное H сравнивается с табличным значением χ^2 (приложение 4) для числа степеней свободы $df = 3 - 1 = 2$. Эмпирическое значение H находится между критическими для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$.

Шаг 6. Принимается статистическое решение и формулируется содержательный вывод. На уровне $\alpha = 0{,}05$ гипотеза H_0 отклоняется. Содержательный вывод: сравниваемые выборки различаются статистически достоверно по уровню выраженности признака ($p < 0{,}05$).

Отметим, что на основании такой проверки мы не можем сделать конкретный вывод о направлении различий и о том, в какой выборке признак принимает большие или меньшие значения. Для этого необходимо парное соотнесение выборок по соответствующему критерию (U-Манна-Уитни).

Обработка на компьютере: критерий *H*-Краскала-Уоллеса

Для обработки использованы данные примера 12.3. В таблице исходных данных (**Data Editor**) для каждого из 16 объектов определены значения двух переменных: `var1` — значения количественного признака, `var2` — группирующая переменная, обозначающая принадлежность каждого объекта к одной из трех сравниваемых групп.

А) Выбираем **Analyze > Nonparametric Tests > K-Independent Samples...** (для *K*-независимых выборок).

Б) В открывшемся окне диалога выделяем и переносим при помощи кнопки ▷ из левого окна интересующие переменные в правое верхнее окно (**Test Variable(s)**) (в данном случае — `var1`); группирующую переменную, которая делит выборку на подгруппы (**Grouping Variable**) (в данном случае — `var2`). Нажимаем кнопку **Define Range...** и задаем диапазон градаций группирующей переменной (градации должны нумероваться подряд) — от минимума (**Minimum**) до максимума (**Maximum**) (в данном случае — от 1 до 3). Нажимаем **Continue**. Нажимаем OK.

В) Получаем результаты в виде двух таблиц:

Ranks

	VAR2	N	Mean Rank
VAR1	1.00	8	5.75
	2.00	5	9.80
	3.00	3	13.67
	Total	16	

Test Statistics(a,b)

	VAR1
Chi-Square	6.575
df	2
Asymp. Sig.	.037

a Kruskal Wallis Test
b Grouping Variable: VAR00006

В первой таблице содержатся ранговые статистики: средние ранги для каждой группы (**Mean Rank**). Во второй таблице содержатся результаты проверки гипотезы: эмпирическое значение критерия χ^2 (**Chi-Square**), число степеней свободы (***df***) и p-уровень значимости (**Asymp. Sig.**).

СРАВНЕНИЕ БОЛЕЕ ДВУХ ЗАВИСИМЫХ ВЫБОРОК

Критерий χ^2-Фридмана (Friedman test) является непараметрическим аналогом однофакторного дисперсионного анализа (ANOVA) для повторных измерений. Он позволяет проверять гипотезы о различии более двух зависимых выборок (повторных измерений) по уровню выраженности изучаемого признака. Критерий χ^2-Фридмана может быть более эффективен, чем его метрический аналог ANOVA в случаях повторных измерений изучаемого признака на небольших выборках.

Критерий χ^2-Фридмана основан на ранжировании ряда повторных измерений для каждого объекта выборки. Затем вычисляется сумма рангов для каждого из условий (повторных измерений). Если выполняется статистическая гипотеза об отсутствии различий между повторными измерениями, то можно ожидать примерное равенство сумм рангов для этих условий. Чем больше различаются зависимые выборки по изучаемому признаку, тем больше эмпирическое значение χ^2-Фридмана.

Эмпирическое значение χ^2-Фридмана вычисляется после ранжирования ряда повторных измерений для каждого объекта по формуле:

$$\chi^2_э = \left[\frac{12}{Nk(k+1)} \cdot \sum_{i=1}^{k} R_i^2\right] - 3N(k+1),\ df = k-1, \qquad (12.3)$$

где N — число объектов (испытуемых), k — количество условий (повторных измерений), R_i — сумма рангов для условия i.

При расчетах «вручную» для определения p-уровня пользуются таблицами критических значений. Если $k = 3$, $N > 9$ или $k > 3$, $N > 4$, то пользуются обычной таблицей для χ^2, $df = k - 1$ (приложение 4). Если $k = 3$, $N < 10$ или $k = 4$,

$N < 5$, то пользуются дополнительными таблицами критических значений χ^2-Фридмана (приложение 13).

При отклонении нулевой статистической гипотезы об отсутствии различий принимается альтернативная гипотеза о статистически достоверных различиях выборок по изучаемому признаку — без конкретизации направления различий. *Для утверждений о том, что уровень выраженности признака в какой-то из сравниваемых выборок выше или ниже, необходимо парное соотнесение выборок по критерию T-Вилкоксона.*

ПРИМЕР 12.4

Проверим гипотезу о различии четырех зависимых выборок по уровню выраженности признака X (о различии четырех условий для одной и той же выборки). Для принятия статистического решения $\alpha = 0{,}05$:

№	Условие 1		Условие 2		Условие 3		Условие 4	
	X	Ранг	*X*	Ранг	*X*	Ранг	*X*	Ранг
1	6	2	14	3.5	5	1	14	3.5
2	11	3	5	2	4	1	12	4
3	12	4	8	2	7	1	10	3
4	8	1	10	2	11	3	12	4
5	5	1	14	3.5	10	2	14	3.5
6	10	3	7	2	6	1	12	4
Сумма рангов:		14		15		9		22

Шаг 1. Для каждого объекта условия ранжируются (по строке).

Шаг 2. Вычисляется сумма рангов для каждого условия: $R_1 = 14$, $R_2 = 15$, $R_3 = 9$, $R_4 = 22$.

Шаг 3. Вычисляется эмпирическое значение χ^2-Фридмана по формуле 12.3:

$$\chi^2_э = \left[\frac{12}{6 \cdot 4 \cdot (4+1)} \cdot (14^2 + 15^2 + 9^2 + 22^2)\right] - 3 \cdot 6 \cdot (4+1) = 8{,}6;\ df = 3.$$

Шаг 4. Определяется p-уровень значимости. Так как $k > 3$, $N > 4$, то пользуются обычной таблицей для χ^2 (приложение 4). Эмпирическое значение χ^2 находится между критическими для $p = 0{,}05$ и $p = 0{,}01$. Следовательно, $p < 0{,}05$.

Шаг 5. Принимается статистическое решение и формулируется содержательный вывод. На уровне $\alpha = 0{,}05$ гипотеза H_0 отклоняется. Содержательный вывод: сравниваемые условия статистически достоверно различаются по уровню выраженности признака ($p < 0{,}05$).

Отметим, что на основании такой проверки мы не можем сделать конкретный вывод о направлении различий и о том, в каких условиях признак принимает большие или меньшие значения. Для этого необходимо парное соотнесение условий по соответствующему критерию (T-Вилкоксона).

Обработка на компьютере: критерий χ^2-Фридмана

Для обработки использованы данные примера 12.4. Исходные данные для обработки введены в таблицу (**Data Editor**) в виде четырех переменных, соответствующих четырем сравниваемым условиям (`var1`, ..., `var4`).

А) Выбираем **Analyze > Nonparametric Tests > K-Related Samples...** (для *k*-зависимых выборок).

Б) В открывшемся окне диалога выделяем переменные (соответствующие нескольким измерениям одного и того же признака) и переносим их при помощи кнопки ▷ из левого окна в правое окно (**Test Variables**). Переменных должно быть больше двух (в данном случае 4). Нажимаем OK.

В) Получаем результаты в виде двух таблиц:

Ranks

	Mean Rank
VAR1	2.33
VAR2	2.50
VAR3	1.50
VAR4	3.67

Test Statistics(a)

N	6
Chi-Square	8.897
df	3
Asymp. Sig.	.031

a Friedman Test

В первой таблице содержатся ранговые статистики: средние ранги для каждой группы (**Mean Rank**). Во второй таблице содержатся результаты проверки гипотезы: эмпирическое значение критерия χ^2 (**Chi-Square**), число степеней свободы (***df***) и *p*-уровень значимости (**Asymp. Sig.**).

Глава 13

ДИСПЕРСИОННЫЙ АНАЛИЗ (ANOVA)

НАЗНАЧЕНИЕ И ОБЩИЕ ПОНЯТИЯ ANOVA[1]

Общепринятое сокращенное обозначение дисперсионного анализа — ANOVA (от англоязычного ANalysis Of VAriance). В соответствии с принятой классификацией, ANOVA — это метод сравнения нескольких (более двух) выборок по признаку, измеренному в метрической шкале. Как и в случае сравнения двух выборок при помощи критерия *t*-Стьюдента, ANOVA решает задачу сравнения средних значений, но не двух, а нескольких. Кроме того, метод допускает сравнение выборок более чем по одному основанию — когда деление на выборки производится по нескольким номинативным переменным, каждая из которых имеет 2 и более градаций.

ПРИМЕР

Исследовалось влияние на продуктивность воспроизведения вербального материала (*Y*): а) интервала между 5 повторениями (X_1 — 3 градации: 1 — 0 мин, 2 — 3 мин, 3 — 10 мин) и б) трудность материала (X_2 — 2 градации: 1 — легкий, 2 — трудный).

Структура данных:

№	X_1 (интервал)	X_2 (объем)	*Y* (эффективность воспроизведения)
1	1	2	8
2	3	2	9
3	2	1	4
4	1	1	5
...	...		...
N	2	2	6

[1] В данной главе содержатся лишь самые необходимые сведения о дисперсионном анализе. Более полное изложение особенностей применения этого мощного и многогранного метода читатель может найти в других источниках, например, в кн.: *Гусева А. Н.* Дисперсионный анализ в экспериментальной психологии. М., 2000.

Специфика ANOVA проявляется в двух отношениях: во-первых, этот метод использует терминологию планирования эксперимента; во-вторых, для сравнения средних значений анализируются компоненты дисперсии изучаемого признака.

ANOVA был разработан Р. Фишером специально для анализа результатов экспериментальных исследований. Соответственно, различные варианты ANOVA воспроизводят наиболее типичные планы организации эксперимента.

Типичная схема эксперимента сводится к изучению влияния независимой переменной (одной или нескольких) на зависимую переменную. *Независимая переменная (Independent Variable)* представляет собой качественно определенный (номинативный) признак, имеющий две или более градации. Каждой градации независимой переменной соответствует выборка объектов (испытуемых), для которых определены значения зависимой переменной. Независимая переменная еще называется *фактором (Factor)*, имеющим несколько градаций (уровней). *Зависимая переменная (Dependent Variable)* в экспериментальном исследовании рассматривается как изменяющаяся под влиянием независимых переменных. В модели ANOVA зависимая переменная должна быть представлена в метрической шкале. В простейшем случае независимая переменная имеет две градации, и тогда задача сводится к сравнению двух выборок по уровню выраженности (средним значениям) зависимой переменной.

В зависимости от соотношения выборок, соответствующих разным градациям (уровням) фактора, различают два типа независимых переменных (факторов). Градациям (уровням) *межгруппового фактора* соответствуют независимые выборки объектов. Градациям (уровням) *внутригруппового фактора* соответствуют зависимые выборки, чаще всего повторные измерения зависимой переменной на одной и той же выборке.

В зависимости от типа экспериментального плана выделяют **четыре основных варианта ANOVA**: однофакторный, многофакторный, ANOVA с повторными измерениями и многомерный ANOVA. Каждый из этих вариантов дисперсионного анализа будет подробно рассмотрен далее в этой главе, а сейчас ограничимся их краткой характеристикой.

Однофакторный ANOVA (One-Way ANOVA) используется при изучении влияния одного фактора на зависимую переменную. При этом проверяется одна гипотеза о влиянии фактора на зависимую переменную.

Многофакторный (двух-, трех-, ... -факторный) ANOVA (2-Way, 3-Way... ANOVA) используется при изучении влияния двух и более независимых переменных (факторов) на зависимую переменную. Многофакторный ANOVA позволяет проверять гипотезы не только о влиянии каждого фактора в отдельности, но и о взаимодействии факторов. Так, для двухфакторного ANOVA проверяются три гипотезы: а) о влиянии одного фактора; б) о влиянии другого фактора; в) о взаимодействии факторов (о зависимости степени влияния одного фактора от градаций другого фактора).

ПРИМЕР

Предположим, изучается влияние на зрительскую оценку различных фильмов (зависимая переменная) двух факторов: жанра фильма (мелодрама, комедия, боевик) и пола зрителя. Вполне вероятно, что в результате такого исследования будут обнаружены не главные эффекты изучаемых факторов (влияние каждого из них в отдельности), а их *взаимодействие*. Взаимодействие факторов «жанр фильма» и «пол зрителя» будет означать, что мужчины и женщины по-разному оценивают фильмы в зависимости от их жанра (фильмы разных жанров оцениваются по-разному, в зависимости от пола зрителя).

ANOVA с повторными измерениями (Repeated Measures ANOVA) применяется, когда по крайней мере один из факторов изменяется по *внутригрупповому плану*, то есть различным градациям этого фактора соответствует одна и та же выборка объектов (испытуемых). Соответственно, в модели ANOVA с повторными измерениями выделяются внутригрупповые и межгрупповые факторы. Для двухфакторного ANOVA с повторными измерениями по одному из факторов проверяются три гипотезы: а) о влиянии внутригруппового фактора; б) о влиянии межгруппового фактора; в) о взаимодействии внутригруппового и межгруппового факторов.

Многомерный ANOVA (Multivariate ANOVA, MANOVA) применяется, когда зависимая переменная является многомерной, иначе говоря, представляет собой несколько (множество) измерений изучаемого явления (свойства).

Дополнительно выделяют *модели ANOVA с постоянными (фиксированными) и случайными эффектами* — различаются способами задания уровней (градаций) фактора. В модели *с постоянными эффектами (Fixed Factors)* уровни остаются фиксированными (одними и теми же) и при проведении данного выборочного исследования: как при обобщении результата на генеральную совокупность, так и при повторном проведении исследования. В модели *со случайными эффектами (Random Factors)* уровни фактора представляют собой более или менее случайную выборку из множества других возможных уровней данного фактора. Конечно, интерпретация (обобщение) результатов будет зависеть от используемой модели. При обработке данных различие между моделями в однофакторном ANOVA может не учитываться, но должно учитываться в двух- (и более) факторном ANOVA. В последнем случае результаты обработки для разных моделей будут различными. Допускается сочетание фиксированных и случайных факторов в одном исследовании.

ПРИМЕР

Сравнивалась эффективность двух учебных программ. Для этого из нескольких сотен школ города было выбрано 5, а в них — по два параллельных класса, ученики которых обучались по разным программам. Исследование представляло собой реализацию двухфакторного плана с одним фиксированным (учебная программа: две градации) и одним случайным факторами (школа: пять градаций). Такое исследование позволяет проверить гипотезу не только об эффективности учебных программ, но и о том, будет ли различаться их эффективность в разных школах города.

В случае, если фактор имеет более двух градаций, то подтверждение гипотезы о его влиянии позволяет сделать лишь неопределенный вывод о том, что по крайней мере две градации фактора различаются. Для более конкретного вывода о том, какие именно градации фактора различаются, в ANOVA предусмотрена процедура *множественных сравнений (Post Hoc multiple comparison tests)*.

Во всех вариантах ANOVA наряду с изучением влияния факторов допускается изучение влияния метрической независимой переменной. Метрическая независимая переменная в этом случае называется *ковариатой* (*Covariate*), и дисперсионный анализ включает в себя *ковариационный анализ.*

Математическая идея ANOVA основана на соотнесении межгрупповой и внутригрупповой частей дисперсии (изменчивости) изучаемой зависимой переменной. Известно, что при объединении двух (или более) выборок с примерно одинаковой дисперсией, но с разными средними значениями дисперсия увеличивается пропорционально различиям средних значений этих выборок. Это связано с тем, что к внутригрупповой дисперсии добавляется дисперсия, обусловленная различиями между группами. В модели ANOVA внутригрупповая изменчивость рассматривается как обусловленная случайными причинами, а межгрупповая — как обусловленная действием изучаемого фактора на зависимую переменную. Соответственно, в общей изменчивости (дисперсии) зависимой переменной выделяются две компоненты: внутригрупповая (случайная) и межгрупповая (факторная) изменчивость. Чем больше отношение межгрупповой изменчивости к внутригрупповой, тем выше факторный эффект — тем больше различаются средние значения, соответствующие разным градациям фактора.

Нулевая гипотеза в ANOVA содержит утверждение о равенстве межгрупповой и внутригрупповой составляющих изменчивости и подразумевает *направленную альтернативу* — о том, что межгрупповая составляющая изменчивости превышает внутригрупповую изменчивость. Нулевой гипотезе соответствует равенство средних значений зависимой переменной на всех уровнях фактора. Принятие альтернативной гипотезы означает, что по крайней мере два средних значения различаются (без уточнения, какие именно градации фактора различаются).

Основные допущения ANOVA: а) распределения зависимой переменной для каждой градации фактора соответствуют нормальному закону; б) дисперсии выборок, соответствующих разным градациям фактора, равны между собой; в) выборки, соответствующие градациям фактора, должны быть независимы (для межгруппового фактора). Выполнение допущения о независимости выборок является обязательным в любом случае. Последствия нарушений остальных двух допущений требуют специального рассмотрения.

Нарушение предположения о нормальности распределения, как показали многочисленные исследования, не оказывает существенного влияния на результаты ANOVA (Шеффе, 1980; Гласс, Стэнли, 1977 и др.). Следовательно, перед проведением ANOVA нет необходимости в проверке соответствия выборочных распределений нормальному закону.

Нарушение предположения о равенстве (однородности, гомогенности) *дисперсий* имеет существенное значение для ANOVA в том случае, *если сравниваемые выборки отличаются по численности.* Таким образом, если выборки, соответствующие разным градациям фактора, отличаются по численности, то необходима предварительная проверка гомогенности (однородности) дисперсий. В компьютерных программах это осуществляется при помощи критерия Ливена (Levene's Test of Homogeneity of Variances). Если выборки заметно различаются по численности и дисперсии по критерию Ливена различаются статистически достоверно, то ANOVA к таким данным не применим, следует воспользоваться непараметрической альтернативой.

В основе современных программных реализаций дисперсионного анализа лежит представление о родственности дисперсионного и множественного регрессионного анализа: оба метода исходят из одной и той же линейной модели. В связи с этим, а также в связи с применением в дисперсионном анализе процедур и показателей, характерных для множественной регрессии, в последнее время все варианты дисперсионного анализа объединяются (например, в программе SPSS) под названием: *Общая линейная модель (GLM — General Linear Model).*

Параметрическими аналогами ANOVA являются такие многомерные методы, как множественный регрессионный анализ (глава 15) и дискриминантный анализ (глава 17). Отличие модели множественного регрессионного анализа заключается в том, что все переменные в ней, в том числе и независимые, представлены в метрической шкале. В модели дискриминантного анализа, в отличие от ANOVA, зависимая переменная является классифицирующей (номинативной), а независимые переменные — метрическими.

Непараметрическими аналогами ANOVA, как отмечалось, являются критерии *H*-Краскала-Уоллеса (для независимых выборок) и χ^2-Фридмана (для повторных измерений).

Вычислительные сложности, связанные с проведением ANOVA, представляли проблему до появления компьютеров и специальных статистических программ. Современные статистические программы (SPSS, STATISTICA) избавляют пользователя от утомительных расчетов. Однако понимание и правильная интерпретация получаемых показателей обязательно требуют наличия общего представления о том, как они вычисляются. Поэтому изложение основных методов ANOVA будет сопровождаться демонстрацией расчетов на упрощенных примерах, которые будущему пользователю компьютерных программ желательно внимательно изучить.

ОДНОФАКТОРНЫЙ ANOVA

Однофакторный (One-Way) ANOVA позволяет проверить гипотезу о том, что изучаемый фактор оказывает влияние на зависимую переменную (средние значения, соответствующие разным градациям фактора, различаются).

Математическая модель однофакторного ANOVA предполагает выделение в общей изменчивости зависимой переменной двух ее составляющих. *Межгрупповая* (факторная) составляющая изменчивости обусловлена различием средних значений под влиянием фактора. *Внутригрупповая* (случайная) составляющая изменчивости обусловлена влиянием неучтенных причин. Соотношение первой и второй из указанных составляющих изменчивости и есть основной показатель, определяющий статистическую значимость влияния фактора (различия средних значений групп, соответствующих уровням фактора).

Нулевая статистическая гипотеза содержит утверждение о равенстве средних значений. При ее отклонении принимается альтернативная гипотеза о том, что по крайней мере два средних значения различаются.

Исходные предположения: распределение зависимой переменной в сравниваемых генеральных совокупностях характеризуется нормальным законом и одинаковыми дисперсиями. Выборки являются случайными и независимыми. Проверка исходных предположений сводится к проверке однородности дисперсии в сравниваемых выборках в случае, если они заметно различаются по численности.

Структура исходных данных: изучаемый признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из нескольких сравниваемых выборок.

ПРИМЕР

Исследовалось влияние на продуктивность воспроизведения (Y) вербального материала интервала между 5 повторениями (X_1 — 3 градации: 1 — 0 мин, 2 — 3 мин, 3 — 10 мин).

Структура данных:

№	X_1 (интервал)	Y (эффективность воспроизведения)
1	1	8
2	3	9
3	2	4
4	1	5
...	...	...
N	2	6

Ограничения: если дисперсии выборок различаются статистически достоверно, то метод неприменим. Для проверки однородности дисперсии применяется критерий Ливена (Levene's Test of Homogeneity of Variances). Формально численность выборок не должна быть менее 2 объектов (фактически необходимо иметь не менее 5 объектов в каждой выборке).

Альтернатива методу: сравнение независимых выборок по критерию H-Краскала-Уоллеса.

Основной результат: принятие или отклонение нулевой статистической гипотезы о равенстве средних значений, соответствующих разным уровням

фактора. Основной показатель для принятия решения — p-уровень значимости критерия F-Фишера.

Дополнительно возможны *множественные сравнения* средних значений, позволяющие сделать вывод о том, как различаются друг от друга средние значения для разных градаций фактора.

Рассмотрим общие принципы и последовательность вычислений для однофакторного ANOVA в случае равной численности сравниваемых выборок.

Исходная идея ANOVA заключается в возможности разложения показателя изменчивости признака на две составляющие: изменчивость внутри групп и изменчивость между группами. В качестве показателя изменчивости используется *сумма квадратов* отклонения значений признака от среднего, которая обозначается SS (Sum of Squares).

Общая (Total) сумма квадратов (SS_{total}) является показателем общей изменчивости зависимой переменной и представляет собой числитель дисперсии:

$$SS_{total} = \sum_{i=1}^{N}(x_i - M)^2 \cdot$$

Соответственно, общая сумма квадратов равна сумме межгрупповой и внутригрупповой сумм квадратов:

$$SS_{total} = SS_{wg} + SS_{bg}.$$

Межгрупповая (Between-Group) сумма квадратов (SS_{bg}) — показатель изменчивости между k группами (каждая численностью n объектов):

$$SS_{bg} = \sum_{j=1}^{k} n(M_j - M)^2 ,$$

где M_j — среднее значение для группы j.

Отношение межгрупповой и общей суммы квадратов показывает *долю общей дисперсии зависимой переменной*, обусловленную влиянием фактора. Этот показатель идентичен по смыслу квадрату коэффициента корреляции в регрессионном анализе, поэтому тоже называется *коэффициентом детерминации (R^2)*:

$$R^2 = \frac{SS_{bg}}{SS_{total}} \cdot$$

Коэффициент детерминации может принимать значения от 0 до 1. Чем больше этот показатель, тем больше влияние изучаемого фактора на дисперсию зависимой переменной. Помноженный на 100, он выражает процент учтенной дисперсии.

Внутригрупповая (Within-Group) сумма квадратов (SS_{wg}) — показатель случайной изменчивости (внутри групп):

$$SS_{wg} = SS_{total} - SS_{bg} = \sum_{j=1}^{k}\sum_{i=1}^{n}(x_i - M_j)^2.$$

На величину сумм квадратов влияет численность и количество сравниваемых групп. Поэтому для сопоставления межгрупповой и внутригрупповой изменчивости используются *средние квадраты* (обозначается MS — от английского Mean of Squares). *Средний квадрат* — это частное от деления суммы квадратов на соответствующее число степеней свободы.

Каждая сумма квадратов характеризуется своим числом степеней свободы (df). Так, *общее число степеней свободы* соответствует общей сумме квадратов и равно:

$$df_{total} = N - 1.$$

Заметим, что частное от деления общей суммы квадратов на общее число степеней свободы — общий средний квадрат — это общая дисперсия.

Число степеней свободы для межгрупповой суммы квадратов равно числу слагаемых минус один (число групп минус 1):

$$df_{bg} = k - 1.$$

Число степеней свободы для внутригрупповой суммы квадратов:

$$df_{wg} = df_{total} - df_{bg} = N - k.$$

После определения числа степеней свободы вычисляются *средние квадраты*:

$$MS_{bg} = \frac{SS_{bg}}{df_{bg}}$$ — межгрупповой средний квадрат;

$$MS_{wg} = \frac{SS_{wg}}{df_{wg}}$$ — внутригрупповой средний квадрат.

Следует отметить, что тот и другой средние квадраты представляют собой различные выборочные оценки одной и той же генеральной дисперсии — для случая, когда сравниваемые средние не различаются. Однако это не так в случае, если хотя бы два из всех сравниваемых средних различаются: тогда межгрупповой средний квадрат превысит внутригрупповой средний квадрат. И чем больше величина отношения межгруппового к внутригрупповому среднему квадрату, тем больше оснований считать, что сравниваемые средние значения различаются. Соответственно, основным показателем ANOVA является *F-отношение* — эмпирическое значение критерия F-Фишера:

$$F_э = \frac{MS_{bg}}{MS_{wg}}, \; df_{bg} = k - 1; \; df_{wg} = N - k.$$

Процедура проверки H_0 подразумевает *направленную альтернативу*, так как ее отклонению соответствует только большее значение $F_э$ ($MS_{bg} > MS_{wg}$). Поэтому для определения p-уровня значимости при вычислениях «вручную» применяются таблицы критических значений F-распределения для направленных альтернатив (односторонний критерий). Для одних и тех же df уровень значимости возрастает (p-уровень убывает) при возрастании $F_э$.

Последовательность выполнения ANOVA является общей для любого числа факторов. Вначале в общей изменчивости зависимой переменной выделяются основные ее составляющие. (В однофакторном ANOVA их две: внутригрупповая (случайная) и межгрупповая (факторная) изменчивость.) После этого вычисляются соответствующие показатели в следующей последовательности:

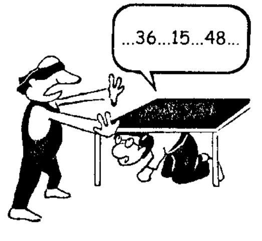
- суммы квадратов (*SS*);
- числа степеней свободы (*df*);
- средние квадраты (*MS*);
- *F*-отношения;
- *p*-уровни значимости.

ПРИМЕР 13.1

Предположим, изучалось различие в продуктивности воспроизведения одного и того же материала трех групп испытуемых (по 5 человек), различающихся условиями предъявления этого материала для запоминания. Зависимая переменная (Y) — количество воспроизведенных единиц материала, независимая переменная (фактор) — условия предъявления (три градации). Проверим на уровне $\alpha = 0{,}01$ гипотезу о том, что продуктивность воспроизведения материала зависит от условий его предъявления.







Так зависит ли запоминание материала от условий его предъявления?

Условие 1		Условие 2		Условие 3	
№	*Y*	№	*Y*	№	*Y*
1	5	6	8	11	11
2	4	7	7	12	9
3	3	8	6	13	7
4	6	9	9	14	10
5	7	10	5	15	8

Общее среднее: $M = 7$.

Среднее для разных условий: $M_1 = 5$; $M_2 = 7$; $M_3 = 9$.

Шаг 1. Вычислим внутригрупповые суммы квадратов:

$$SS_{total} = \sum_{i=1}^{N}(x_i - M)^2 = (5-7)^2 + (4-7)^2 + \ldots + (8-7)^2 = 70;$$

$$SS_{bg} = \sum_{j=1}^{k} n(M_j - M)^2 = 5\left[(5-7)^2 + (7-7)^2 + (9-7)^2\right] = 40;$$

$$SS_{wg} = SS_{total} - SS_{bg} = 70 - 40 = 30.$$

Шаг 2. Определим числа степеней свободы:

$$df_{bg} = k - 1 = 3 - 1 = 2; \; df_{wg} = N - k = 15 - 3 = 12.$$

Шаг 3. Вычислим средние квадраты:

$$MS_{bg} = \frac{SS_{bg}}{df_{bg}} = \frac{40}{2} = 20; \; MS_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{30}{12} = 2{,}5.$$

Шаг 4. Вычислим F-отношение:

$$F_э = \frac{MS_{bg}}{MS_{wg}} = \frac{20}{2{,}5} = 8.$$

Шаг 5. Определим p-уровень значимости. По таблице критических значений F-распределения (для направленных альтернатив) (приложение 3) для $p = 0{,}01$; $df_{числ} = 2$; $df_{знам} = 12$ критическое значение равно $F = 6{,}927$. Следовательно, $p < 0{,}01$. Дополнительно вычислим коэффициент детерминации:

$$R^2 = \frac{SS_{bg}}{SS_{total}} = \frac{40}{70} = 0{,}571.$$

Оформим результат в виде таблицы:

Источник изменчивости	Сумма квадратов (SS)	df	Средний квадрат (MS)	F	p-уровень
Межгрупповой	40	2	20	8	< 0,01
Внутригрупповой	30	12	2,5	—	—
Общий	70	14	—	—	—

$R^2 = 0{,}571$.

Шаг 6. Принимаем статистическое решение и формулируем содержательный вывод. Отклоняем H_0 и принимаем альтернативную гипотезу о том, что межгрупповая изменчивость выше внутригрупповой. Содержательный вывод: обнаружено статистически достоверное влияние условий предъявления материала на продуктивность его воспроизведения ($p < 0{,}01$). Или: средние значения продуктивности воспроизведения материала статистически достоверно различаются в зависимости от условий его предъявления ($p < 0{,}01$).

В ANOVA *с выборками неравной численности* вычисления несколько усложняются. Изменения касаются формулы для вычисления межгрупповой суммы квадратов:

$$SS_{bg} = \sum_{j=1}^{k} n_j (M_j - M)^2,$$

где n_j — численность группы j.

Кроме того, если группы различаются по численности, необходима проверка гомогенности дисперсии с использованием критерия Ливена.

Обработка на компьютере

Рассмотрим применение однофакторного ANOVA к данным примера 13.1. Исходные данные для анализа введены в таблицу (**Data Editor**) в следующем виде:

	vospr	f1
1	5	1
2	4	1
3	3	1
4	6	1
5	7	1
6	8	2
7	7	2
8	6	2
9	9	2
10	5	2
11	11	3
12	9	3
13	7	3
14	10	3
15	8	3

1. Выбираем **Analyze > Compare means > One Way ANOVA...**

2. В открывшемся окне диалога выделяем и переносим из левого окна переменные при помощи кнопки ▷: зависимую переменную (vospr) в правое верхнее окно (**Dependent List**), переменную, соответствующую фактору (f1), — в правое нижнее окно (**Factor**). Нажимаем **Options...** В открывшемся окне диалога отмечаем флажком: **Descriptive** (Описательные статистики), **Homogeneity of variance test** (Тест однородности дисперсии), **Means plot** (График средних значений). Нажимаем **Continue** (Продолжить). Нажимаем OK.

3. Получаем результаты.

А) Описательные статистики:

Descriptives
VOSPR

	N	Mean	Std. Deviation	Std. Error
1.00	5	5.0000	1.58114	.70711
2.00	5	7.0000	1.58114	.70711
3.00	5	9.0000	1.58114	.70711
Total	15	7.0000	2.23607	.57735

Первая колонка — номера градаций фактора, вторая колонка (N) — численность выборок, Mean — средние значения, Std. Deviation — стандартное отклонение, Std. Error — ошибка среднего.

B) Проверка однородности дисперсии:

Test of Homogeneity of Variances
VOSPR

Levene Statistic	df1	df2	Sig.
.000	2	12	1.000

Тест Ливена показывает, что статистически достоверных различий между дисперсиями не обнаружено (Sig. > 0,05). Следовательно, допустимо применение ANOVA.

C) Результаты ANOVA:

ANOVA
VOSPR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	40.000	2	20.000	8.000	.006
Within Groups	30.000	12	2.500		
Total	70.000	14			

Результаты соответствуют тем, которые были получены при обработке этих данных «вручную»: условия предъявления слов статистически достоверно влияют на продуктивность их воспроизведения.

D) График средних значений:

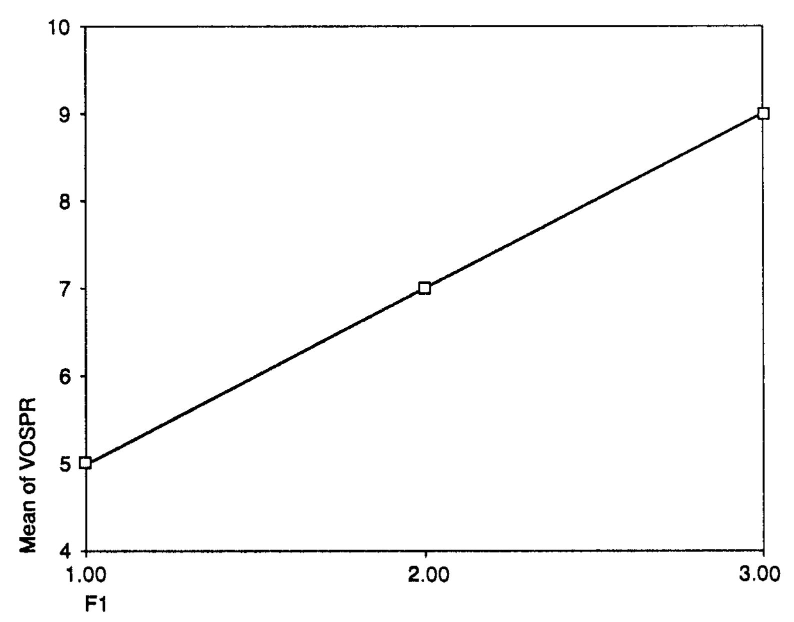


График средних значений облегчает интерпретацию факторного эффекта: продуктивность воспроизведения монотонно возрастает от первого к третьему условию предъявления.

МНОЖЕСТВЕННЫЕ СРАВНЕНИЯ В ANOVA

В состав процедур ANOVA включаются множественные сравнения средних значений для разных уровней фактора: парные сравнения средних после отклонения H_0(Post Hoc Tests) ; метод контрастов (Contrasts).

Методы сравнения средних после отклонения H_0 об отсутствии различий предназначены для выделения тех пар средних, которые привели к отклонению H_0. Эти методы сводятся к последовательному сопоставлению всех пар средних значений для одного фактора. Применение для этих целей, казалось бы, подходящего критерия t-Стьюдента является некорректным, так как дело касается проверки одновременно нескольких гипотез. Тем не менее, разработано множество процедур корректного множественного сравнения пар средних (методы Бонферрони, Тьюки, Дункан, Шеффе и др.). Рассмотрим один из них — наиболее популярный *метод Шеффе (Scheffé test).*

При использовании метода Шеффе достоверность различия средних значений определяется по формуле эмпирического значения критерия t-Шеффе:

$$t_э = \frac{M_1 - M_2}{\sqrt{\frac{MS_{wg}}{n_1} + \frac{MS_{wg}}{n_2}}},$$

где M_1, M_2 — сравниваемые средние значения; n_1, n_2 — численность соответствующих групп; MS_{wg} — внутригрупповой средний квадрат. Для определения p-уровня эмпирическое значение сравнивается с критическим значением, которое в свою очередь вычисляется по формуле исходя из критического значения F-критерия для df_{bg} и df_{wg}:

$$t_{кр} = \sqrt{F_{кр}(k-1)} .$$

Ограничение на применение метода Шеффе: дисперсии в сравниваемых выборках, соответствующих уровням фактора, не должны статистически достоверно различаться. Для проверки однородности дисперсии применяется критерий Ливена (Levene's Test of Homogeneity of Variances). Если дисперсии различаются, то следует воспользоваться другими критериями, которые предлагает для этого случая компьютерная программа (SPSS): **Tamhane's T2, Dunnett's T3, Games-Howell, Dunnett's C**.

ПРИМЕР 13.2

Сравним уровни фактора для предыдущего примера 13.1.

Шаг 1. Вычислим эмпирические значения критерия t-Шеффе:

$$t_{12} = \frac{|5-7|}{\sqrt{\frac{2,5}{5} + \frac{2,5}{5}}} = 2; \; t_{13} = \frac{|5-9|}{\sqrt{\frac{2,5}{5} + \frac{2,5}{5}}} = 4; \; t_{23} = \frac{|7-9|}{\sqrt{\frac{2,5}{5} + \frac{2,5}{5}}} = 2 .$$

Ш а г 2. Вычислим критическое значение t-Шеффе.

Для $p = 0{,}01$; $df_{bg} = 2$; $df_{wg} = 12$; $F_{кр} = 6{,}927$; $t_{кр} = \sqrt{F_{кр}(k-1)} = \sqrt{6{,}927 \cdot 2} = 3{,}722$.

Ш а г 3. Определяем p-уровень значимости для каждой пары средних значений по таблице критических значений t-критерия (приложение 2):

$$p_{12} > 0{,}05;\ p_{13} < 0{,}01;\ p_{23} > 0{,}05.$$

Ш а г 4. Принимаем статистические решения и формулируем содержательный вывод. Гипотеза о равенстве средних значений отклоняется только для уровней 1 и 3. Влияние условий предъявления материала на продуктивность его воспроизведения проявляется в статистически достоверном различии условий 1 и 3: средняя продуктивность воспроизведения при условии 3 выше, чем при условии 1 ($p < 0{,}01$).

Метод контрастов (Contrasts) не предполагает обязательного отклонения H_0 и позволяет оценить различия между *сочетаниями* средних значений для разных уровней фактора. Например, можно сравнить общее среднее значение первого и второго уровней со средним значением для третьего уровня фактора. *Контраст* (K) — это линейная комбинация сравниваемых средних значений, которая задается в виде *полинома*:

$$K = c_1M_1 + c_2M_2 + \ldots + c_kM_k,$$
$$\text{такая, что } c_1 + c_2 + \ldots + c_k = 0.$$

ПРИМЕР

Если фактор имеет три градации и нас интересует отличие первой градации от двух других, то контрастом будет выражение:

$$K = 2M_1 - M_2 - M_3, \text{ или } K = M_1 - 0{,}5M_2 - 0{,}5M_3$$

а коэффициенты контраста: $2 - 1 - 1 = 0$, или $1 - 0{,}5 - 0{,}5 = 0$.

Таким образом, задав вид полинома, можно оценить соотношение между средними значениями (при игнорировании какого-либо уровня ему присваивается коэффициент 0).

Проверка достоверности отличия контраста от нуля производится по формуле эмпирического значения критерия t-Шеффе:

$$t_э = \frac{K}{\sqrt{MS_{wg}\left(\frac{c_1^2}{n_1} + \frac{c_2^2}{n_2} + \ldots + \frac{c_k^2}{n_k}\right)}},$$

где $K = c_1M_1 + c_2M_2 + \ldots + c_kM_k$, $c_1 + c_2 + \ldots + c_k = 0$.

Для определения p-уровня эмпирическое значение сравнивается с критическим значением двустороннего t-распределения (для ненаправленных альтернатив) для $df = df_{wg} = N - k$[1].

[1] Так определяется p-уровень в программе SPSS. В других источниках предлагается более консервативный метод — вычисление критического значения по формуле Шеффе (Гласс, Стэнли, 1977), — увеличивающий значение p-уровня.

Ограничение на применение метода контрастов: дисперсии в сравниваемых выборках не должны статистически достоверно различаться. Для проверки однородности дисперсии применяется критерий Ливена (Levene's Test of Homogeneity of Variances). При различии дисперсий компьютерные программы (SPSS) вводят поправку в число степеней свободы и, соответственно, корректируют p-уровень значимости.

ПРИМЕР 13.3

Определим для примера 13.1 достоверность отличия уровней 1 и 2 от уровня 3.

Шаг 1. Зададим коэффициенты контраста: $c_1 = 1$; $c_2 = 1$; $c_3 = -2$.

Шаг 2. Определим эмпирическое значение критерия t-Шеффе:

$$K = M_1 + M_2 - 2M_3 = 5 + 7 - 18 = -6,$$

$$t_э = \frac{|5+7-18|}{\sqrt{2{,}5\left(\frac{1}{5}+\frac{1}{5}+\frac{4}{5}\right)}} = \frac{6}{\sqrt{2{,}5 \cdot 1{,}2}} = 3{,}464\,.$$

Отметим, что если задать коэффициенты контраста 0,5; 0,5; –1 ($K = -3$), то величина $t_э$ не изменится.

Шаг 3. Определяем p-уровень, сопоставляя эмпирическое значение с табличными критическими значениями t-распределения (приложение 2) для $df_{wg} = 12$: $p < 0{,}01$.

Шаг 4. Принимаем статистическое решение и формулируем вывод. Контраст статистически достоверно отличается от нуля. Продуктивность воспроизведения при условии 3 статистически достоверно выше, чем средняя продуктивность воспроизведения для условий 1 и 2 ($p < 0{,}01$).

Обработка на компьютере

Рассмотрим применение методов множественного сравнения с использованием данных примера 13.1. Применим метод Шеффе для парного сравнения средних и метод контрастов для сравнения третьего уровня фактора с двумя другими его уровнями.

Повторим все операции, которые мы совершали для проведения однофакторного ANOVA:

1. Выбираем **Analyze > Compare means > One Way ANOVA...**

2. В открывшемся окне диалога выделяем и переносим из левого окна переменные при помощи кнопки ▷: зависимую переменную (`prod`) в правое верхнее окно (**Dependent List**); переменную, соответствующую фактору (`f1`), — в правое нижнее окно (**Factor**). Нажимаем **Options...** В открывшемся окне диалога отмечаем флажком: **Descriptive** (Описательные статистики), **Homogeneity of variance test** (Тест однородности дисперсии), **Means plot** (График средних значений). Нажимаем **Continue** (Продолжить).

Для *парного сравнения средних* в окне диалога **One way ANOVA** дополнительно нажимаем кнопку **Post Hoc...** (Постфактум, то есть после отклонения H_0). В открывшемся окне диалога отмечаем флажком необходимый нам метод сравнения: **Scheffe** (Шеффе) (при желании можно было бы выбрать и другие методы, в частности те, применение которых не требует однородности дисперсии сравниваемых выборок). Нажимаем **Continue** (Продолжить).

Для применения *метода контрастов* в окне диалога **One way ANOVA** дополнительно нажимаем кнопку **Contrasts...** (Контрасты...). В открывшемся окне диалога отмечаем флажком **Polynomial** (Полином) и последовательно задаем коэффициенты полинома для контраста. Последовательность коэффициентов должна соответствовать последовательности уровней фактора (от меньшего к большему). Сумма коэффициентов должна быть равна 0. Вводим в окне **Coefficients** (Коэффициенты) сначала 1, нажимаем **Add** (Добавить), затем **1**, снова **Add**, затем **—2** и **Add**. В окне ниже увидим значения коэффициентов и ниже — их сумму (**Coefficient Total: 0.00**). Если сумма равна 0, значит коэффициенты назначены верно. После этого можно составить другой контраст, для чего следует нажать клавишу **Next** (Следующий). После назначения контрастов нажимаем **Continue** (Продолжить). Нажимаем OK.

3. Получаем результаты.

Дополнительно к тем результатам, которые были описаны для одномерного ANOVA, получим следующие результаты:

А) Коэффициенты контраста:

Contrast Coefficients

Contrast	F1		
	1.00	2.00	3.00
1	1	1	-2

В) Результаты статистической проверки контраста:

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig.(2-tailed)
VOSPR	Assume equal variances	1	-6.0000	1.73205	-3.464	12	.005
	Does not assume equal variances	1	-6.0000	1.73205	-3.464	8.000	.009

Столбец `Contrast` показывает номер контраста (1): их будет столько, сколько было введено (в данном случае он один). `Value of Contrast` (Значение контраста) — разность, статистическая значимость которой проверяется. `Std. Error` — стандартная ошибка контраста. `t` — значение *t*-критерия, `df` — число степеней свободы, `Sig.` — *p*-уровень значимости контраста. Первая строчка таблицы дает результаты контраста для случая, когда диспер-

сии сравниваемых групп (уровней) однородны, а вторая — для случая неоднородности дисперсий по критерию Ливена.

Получены те же результаты, что и при вычислении «вручную» (пример 13.3). По результатам можно сделать вывод о статистически достоверно более высокой продуктивности воспроизведения слов при третьем условии, по сравнению с двумя другими условиями.

C) Результаты парных сравнений средних значений по методу Шеффе:

Post Hoc Tests
Multiple Comparisons
Dependent Variable: VOSPR
Scheffe

(I) F1	(J) F1	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1.00	2.00	-2.0000	1.00000	.178	-4.7876	.7876
	3.00	-4.0000(*)	1.00000	.006	-6.7876	-1.2124
2.00	1.00	2.0000	1.00000	.178	-.7876	4.7876
	3.00	-2.0000	1.00000	.178	-4.7876	.7876
3.00	1.00	4.0000(*)	1.00000	.006	1.2124	6.7876
	2.00	2.0000	1.00000	.178	-.7876	4.7876

* The mean difference is significant at the .05 level.

Так же, как и для вычислений «вручную» (пример 13.2), получено статистически значимое различие между уровнями 1 и 3 (Sig. = 0,006).

Дополнительно выдаются результаты проверки однородности дисперсии для сравниваемых выборок:

Homogeneous Subsets
VOSPR
Scheffe

F1	N	Subset for alpha = .05	
		1	2
1.00	5	5.0000	
2.00	5	7.0000	7.0000
3.00	5		9.0000
Sig.		.178	.178

Means for groups in homogeneous subsets are displayed.
a Uses Harmonic Mean Sample Size = 5.000.

Результаты демонстрируют отсутствие статистически достоверных различий дисперсий 1 и 2 (Sig. = 0,178), 2 и 3 (Sig. = 0,178) выборок, что убеждает в корректности парных сравнений средних значений.

МНОГОФАКТОРНЫЙ ANOVA

Многофакторный ANOVA предназначен для изучения влияния нескольких факторов (независимых переменных) на зависимую переменную и часто обозначается в соответствии с количеством факторов и числом их градаций. Например, обозначение ANOVA 3×2×2 свидетельствует о трехфакторном ANOVA (число градаций: первого фактора — 3, второго фактора — 2, третьего фактора — 2), который применяется для сравнения 12 групп (условий) (так как 3×2×2 = 12).

Принципиально этот метод не отличается от однофакторного ANOVA. Однако он позволяет оценивать не только влияние (главные эффекты) каждого фактора в отдельности, но и *взаимодействие факторов*: зависимость влияния одних факторов от уровней других факторов. Возможность изучать взаимодействие факторов — главное преимущество многофакторного ANOVA, которое позволяет получать зачастую наиболее интересные результаты исследования.

С целью облегчения изложения материала в качестве основного варианта многофакторного ANOVA мы сначала рассмотрим двухфакторный его вариант (2-Way ANOVA), а затем сделаем необходимые дополнения в отношении большего количества факторов.

Структура исходных данных (2-факторный ANOVA). Для каждого объекта (испытуемого) выборки измерено значение зависимой переменной (Y), а также определена его принадлежность к одной из градаций (уровней) одного фактора (X_1) и к одной из градаций (уровней) другого фактора (X_2). Таблица исходных данных для компьютерной обработки включает две номинативные переменные, соответствующие факторам, и одну метрическую (зависимую) переменную:

№ объектов	X_1 (Фактор 1)	X_2 (Фактор 2)	Y (Зависимая переменная)
1	1	2	8
2	3	2	9
3	2	1	4
4	1	1	5
...	...		...
N	2	2	6

Модель для данных может быть представлена в виде *дисперсионного комплекса* — таблицы, строки которой соответствуют градациям (уровням) одного фактора: 1, 2, ..., i, ..., k; а столбцы — уровням другого фактора: 1, 2, ..., j, ..., l. Количество *ячеек* дисперсионного комплекса равно $k \times l$ и соответствует количеству разных групп объектов (испытуемых). Каждая ячейка с номером ij характеризуется своим сочетанием уровней факторов, численностью объектов n_{ij} и средним значением зависимой переменной M_{ij}. Например, дисперсионный комплекс для ANOVA 2×3:

Фактор *A*	Фактор *B*			
	1	2	3	
1	M_{11}	M_{12}	M_{13}	M_{A1}
2	M_{21}	M_{22}	M_{23}	M_{A2}
	M_{B1}	M_{B2}	M_{B3}	M

Математическая модель двухфакторного ANOVA, как и в однофакторном случае, предполагает выделение двух основных частей вариации зависимой переменной: внутригрупповой, обусловленной случайными причинами, и межгрупповой, обусловленной влиянием факторов. В межгрупповой изменчивости, в свою очередь, выделяются три ее составляющие:

- влияние (главный эффект) 1-го фактора;
- влияние (главный эффект) 2-го фактора;
- взаимодействие факторов.

Соответственно, двухфакторный ANOVA включает в себя проверку трех гипотез: а) о главном эффекте 1-го фактора; б) о главном эффекте 2-го фактора; в) о взаимодействии факторов.

Проблема взаимодействия факторов, которая обеспечивает уникальность и незаменимость многофакторного ANOVA, заслуживает отдельного рассмотрения. Понятие взаимодействия двух независимых факторов было введено основателем дисперсионного анализа Р. Фишером для обозначения ситуации, когда влияние одного фактора на зависимую переменную проявляется *по-разному* на разных уровнях другого фактора.

ПРИМЕР 13.4 (Солсо Р., МакЛин М. К., с. 58–59)

Студентам колледжа предложили написать сочинение в поддержку закона о самоуправлении, противниками которого все они являлись. Испытуемым либо давали задание написать такое сочинение (условие без выбора), либо предлагали самим выбирать — писать или не писать (условие с выбором) (фактор *A*: 2 уровня). Кроме того, половине испытуемых в каждой из групп платили по 0,5$, а другой половине — 2,5$ за написание этого сочинения (фактор *B*: 2 уровня). В каждую из 4-х групп случайно отбиралось по 10 студентов. Зависимой переменной являлась степень изменения отношения студентов к закону о самоуправлении после написания сочинения. Средние значения изменения отношения для различных групп:

Фактор *A*	Фактор *B*		Средние
	0,5$ (1)	2,5$ (2)	
Нет выбора (1)	–0,05	+0,63	0,29
Свободный выбор (2)	+1,25	–0,07	0,59
Средние:	0,6	0,28	0,44

Результаты (рис. 13.1) демонстрируют взаимодействие факторов: размер вознаграждения (фактор *B*) по-разному влияет на изменение отношения — в зависимости от наличия или отсутствия свободного выбора (фактор *A*).

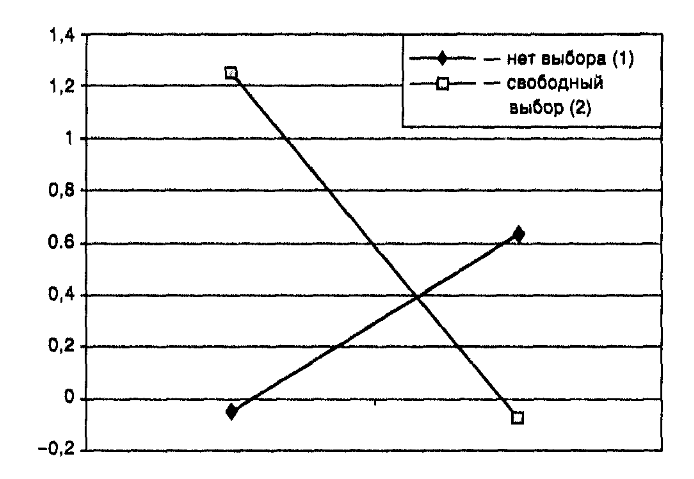
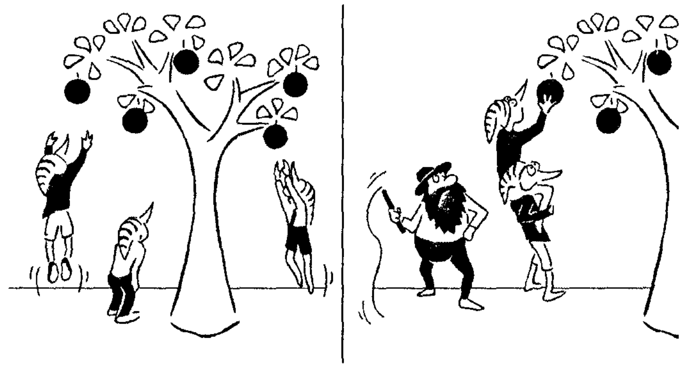


Рис. 13.1. График средних значений изменения отношения к закону о самоуправлении (к данным примера 13.4)

В условиях отсутствия выбора отношение испытуемых к закону о самоуправлении улучшилось в случае большего вознаграждения; в условиях же свободного выбора наблюдалась обратная картина: более хорошее отношение продемонстрировали те, кто получил меньшее вознаграждение.

ПРИМЕР 13.5

Предположим, изучается влияние на успешность группового решения задачи численности группы и наличия или отсутствия лидера в группе. Зависимая переменная — время решения задачи в минутах. Фактор А — размер группы, три градации: 1 — 2–3 человека; 2 — 5–7 человек; 3 — 10–15 человек. Фактор В — наличие лидера: 1 — есть; 2 — нет. В качестве объектов выступают группы. В зависимости от стиля лидерства, сложности задания и других причин, которые не учитываются, можно было бы получить разные эффекты взаимодействия факторов численности группы и наличия лидерства (рис. 13.2). График 1 демонстрирует сильное взаимодействие факторов (группы большей численности более эффективны, если в них есть лидер, а группы малой численности — при отсутствии лидера), а график 3 — более слабое взаимодействие (наличие лидера играет роль лишь в группах большой численности). Графики 2 и 4 соответствуют ситуации отсутствия взаимодействия.



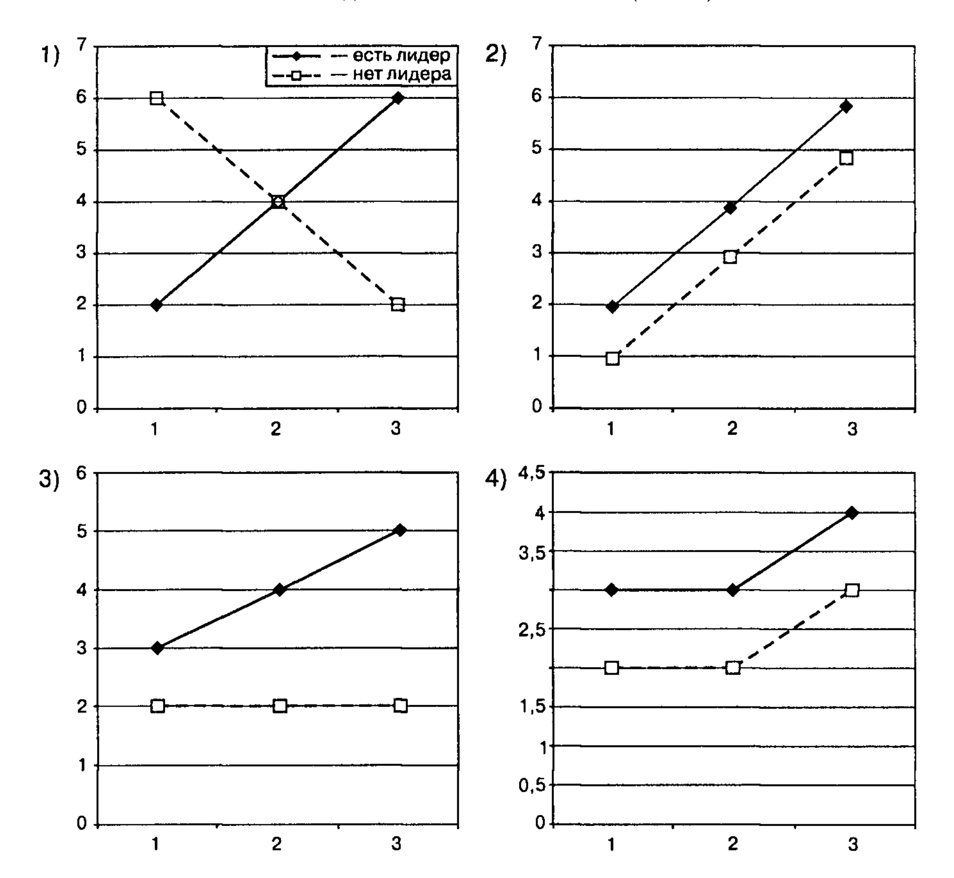


Рис. 13.2. Графики средних значений успешности группового решения задачи (к данным примера 13.5)

Приведенные примеры демонстрируют эффективность визуального анализа графиков средних значений: если линии, соответствующие разным уровням одного из факторов, не параллельны, то можно предполагать наличие взаимодействия факторов. Однако окончательное заключение об этом можно сделать только при статистическом подтверждении гипотезы о взаимодействии по результатам ANOVA. Таким образом, графики средних значений особенно полезны для интерпретации обнаруженного статистически достоверного взаимодействия факторов.

Исходные предположения многофакторного ANOVA: распределение зависимой переменной в сравниваемых генеральных совокупностях (соответствующих ячейкам дисперсионнго комплекса) характеризуется нормальным законом и одинаковыми дисперсиями. Выборки в каждой ячейке являются случайными и независимыми.

Ограничения: если выборки (ячейки) заметно различаются по численности и их дисперсии различаются статистически достоверно, то метод неприменим. Число наблюдений в каждой ячейке не должно быть меньше 2 (желательно — не менее 5). Проверка допустимости применения ANOVA сводится к про-

верке однородности дисперсии в сравниваемых выборках в случае, если они заметно различаются по численности. Для проверки однородности дисперсии применяется критерий Ливена (Levene's Test of Homogeneity of Variances).

Дополнительно возможны множественные сравнения средних значений, позволяющие сделать вывод о том, как различаются друг от друга средние значения, соответствующие разным градациям факторов.

Общая схема двух- (и более) факторного ANOVA принципиально не отличается от однофакторного случая и определяется выделением в общей изменчивости зависимой переменной (SS_{tot}) ее внутригрупповой (случайной, SS_{wg}) и межгрупповой (факторной, SS_{bg}) составляющих:

$$SS_{tot} = SS_{wg} + SS_{bg}.$$

Отличие заключается в выделении дополнительных составляющих межгрупповой (факторной) изменчивости в соответствии с проверяемыми гипотезами. Для двухфакторного случая:

$$SS_{bg} = SS_A + SS_B + SS_{AB},$$

где SS_A, SS_B – суммы квадратов для факторов A и B, а SS_{AB} – сумма квадратов для взаимодействия факторов. Соответственно, для каждого источника изменчивости далее вычисляются степени свободы и средние квадраты, вычисляются F-отношения для проверяемых гипотез и определяются p-уровни значимости.

Последовательность вычислений основных показателей для двухфакторного ANOVA рассмотрим на упрощенном примере — при равной численности сравниваемых выборок (объектов в ячейках). Для случая с неравной численностью наблюдений в ячейках логика и общая последовательность вычислений не меняются, хотя сами вычисления и становятся более громоздкими.

Фактор A	Фактор B			
	1	2	3	
1	M_{11}	M_{12}	M_{13}	M_{A1}
2	M_{21}	M_{22}	M_{23}	M_{A2}
	M_{B1}	M_{B2}	M_{B3}	M

Численность каждой ячейки равна n, общее число наблюдений — $6n = N$.

Напомним, что двухфакторный ANOVA проверяет 3 статистические гипотезы: а) о главном эффекте фактора A (о различии M_{A1} и M_{A2}); б) о главном эффекте фактора B (о различии M_{B1}, M_{B2} и M_{B3}); в) о взаимодействии факторов A и B (влияние фактора A различается для разных уровней фактора B, и наоборот).

Межгрупповая (SS_{bg}) и *внутригрупповая* (SS_{wg}) *суммы квадратов* вычисляются как составные части общей суммы квадратов (SS_{tot}):

$$SS_{tot} = \sum_{i=1}^{N}(x_i - M)^2 .$$

$$SS_{bg}=\sum_{i=1}^{k}\sum_{j=1}^{l}n(M_{ij}-M)^2\ ,\ SS_{wg}=SS_{total}-SS_{bg},$$

где k — число уровней фактора A; l — число уровней фактора B; M_{ij} — среднее значение для ячейки ij.

Отношение межгрупповой и общей суммы квадратов — коэффициент детерминации. Как и в однофакторном случае, он показывает долю общей дисперсии зависимой переменной, которая обусловлена совокупным влиянием факторов (факторной моделью):

$$R^2=\frac{SS_{bg}}{SS_{total}}.$$

Чем больше этот показатель, тем больше общая дисперсия зависимой переменной объясняется влиянием изучаемых факторов. Межгрупповая сумма квадратов состоит из трех составляющих ее сумм квадратов: для фактора A, для фактора B, для взаимодействия факторов A и B:

$$SS_{bg}=SS_A+SS_B+SS_{AB}.$$

Суммы квадратов для фактора A (SS_A) и фактора B (SS_B):

$$SS_A=ln\sum_{i=1}^{k}\left[\frac{1}{l}\sum_{j=1}^{l}M_{ij}-M\right]^2=3n\left[(M_{A1}-M)^2+(M_{A2}-M)^2\right],$$

$$SS_B=kn\sum_{i=1}^{l}\left[\frac{1}{k}\sum_{j=1}^{k}M_{ij}-M\right]^2=2n\left[(M_{B1}-M)^2+(M_{B2}-M)^2+(M_{B3}-M)^2\right].$$

Сумма квадратов для взаимодействия факторов A и B — это остаток межгрупповой суммы квадратов за вычетом сумм квадратов факторов A и B:

$$SS_{AB}=SS_{bg}-SS_A-SS_B.$$

Числа степеней свободы для сумм квадратов:
для общей: $df_{tot}=N-1$;
для фактора A: $df_A=k-1$;
для фактора B: $df_B=l-1$;
для взаимодействия факторов: $df_{AB}=df_A\times df_B$;
для внутригрупповой: $df_{wg}=df_{tot}-df_A-df_B-df_{AB}=N-k\times l$;
для общей межгрупповой (факторной): $df_{bg}=k\times l-1$.

Средние квадраты вычисляются делением сумм квадратов на соответствующие им числа степеней свободы:

$$MS_A=\frac{SS_A}{df_A};\ MS_B=\frac{SS_B}{df_B};\ MS_{AB}=\frac{SS_{AB}}{df_{AB}};\ MS_{wg}=\frac{SS_{wg}}{df_{wg}}.$$

Вычисляются *эмпирические значения F-отношения* для каждой из трех проверяемых гипотез:

$$F_A = \frac{MS_A}{MS_{wg}};\ F_B = \frac{MS_B}{MS_{wg}};\ F_{AB} = \frac{MS_{AB}}{MS_{wg}}.$$

Дополнительно можно вычислить F-отношение для общей факторной модели, которое позволит определить статистическую значимость совокупного влияния факторов:

$$F_{bg} = \frac{MS_{bg}}{MS_{wg}}.$$

Для определения p-уровня значимости каждого из F-отношения вычисленное эмпирическое значение сравнивается с критическими (табличными) значениями для степеней свободы, соответствующих числителю и знаменателю F-отношения.

ПРИМЕР 13.6

Предположим, изучается влияние численности группы и наличия или отсутствия лидера в группе на успешность группового решения задачи. В одной из серий исследования получены следующие результаты:

Время решения тестовой задачи группами разной численности в зависимости от наличия или отсутствия лидера

Группы без лидера:			Группы с лидером:		
Малая (1)	Средняя (2)	Большая (3)	Малая (1)	Средняя (2)	Большая (3)
4	9	8	9	10	7
8	5	8	11	7	5
5	7	9	10	8	4
7	6	6	8	8	6
6	8	9	12	7	8
$M_{11} = 6$	$M_{12} = 7$	$M_{13} = 8$	$M_{21} = 10$	$M_{22} = 8$	$M_{23} = 6$

В качестве объектов выступают группы. Зависимая переменная — время решения задачи в минутах. Фактор A — наличие лидера: 1 — нет; 2 — есть. Фактор B — размер группы, три градации: 1 — 2–3 человека; 2 — 5–7 человек; 3 — 10–15 человек. Проверим гипотезы о влиянии факторов и их взаимодействия на уровне $\alpha = 0{,}05$.

Шаг 1. Составим дисперсионный комплекс и подсчитаем средние значения:

Фактор A	Фактор B			
	1	2	3	
1	$M_{11} = 6$	$M_{12} = 7$	$M_{13} = 8$	$M_{A1} = 7$
2	$M_{21} = 10$	$M_{22} = 8$	$M_{23} = 6$	$M_{A2} = 8$
	$M_{B1} = 8$	$M_{B2} = 7{,}5$	$M_{B3} = 7$	$M = 7{,}5$

Шаг 2. Вычислим межгрупповую (SS_{bg}) и внутригрупповую (SS_{wg}) суммы квадратов как составные части общей суммы квадратов (SS_{tot}):

$$SS_{tot} = \sum_{i=1}^{30}(x_i - M)^2 = 109{,}5,$$

$$SS_{bg} = \sum_{i=1}^{2}\sum_{j=1}^{3} 5(M_{ij} - M)^2 = 5\left[(6-7{,}5)^2 + (7-7{,}5)^2 + \ldots\right] = 57{,}5\,,$$

$$SS_{wg} = SS_{total} - SS_{bg} = 109{,}5 - 57{,}5 = 52.$$

Доля общей изменчивости, объясняемая данной факторной моделью:

$$R^2 = \frac{SS_{bg}}{SS_{tot}} = \frac{57{,}5}{109{,}5} = 0{,}525.$$

Ш а г 3. Вычислим суммы квадратов для фактора A (SS_A), фактора B (SS_B) и взаимодействия факторов (SS_{AB}):

$$SS_A = 3n\left[(M_{A1} - M)^2 + (M_{A2} - M)^2\right] = 15\left[(7-7{,}5)^2 + (8-7{,}5)^2\right] = 7{,}5$$

$$SS_B = 2n\left[(M_{B1} - M)^2 + (M_{B2} - M)^2 + (M_{B3} - M)^2\right] = 10\left[(8-7{,}5)^2 + (7{,}5-7{,}5)^2 + (7-7{,}5)^2\right] = 5;$$

$$SS_{AB} = SS_{bg} - SS_A - SS_B = 57{,}5 - 7{,}5 - 5 = 45.$$

Ш а г 4. Определим степени свободы для вычисленных сумм квадратов:

для общей: $df_{tot} = N - 1 = 30 - 1 = 29$;

для фактора A: $df_A = k - 1 = 2 - 1 = 1$;

для фактора B: $df_B = l - 1 = 3 - 1 = 2$;

для взаимодействия факторов: $df_{AB} = df_A \cdot df_B = 1 \cdot 2 = 2$;

для внутригрупповой: $df_{wg} = df_{tot} - df_A - df_B - df_{AB} = N - k \cdot l = 30 - 6 = 24$;

для общей межгрупповой (факторной): $df_{bg} = k \cdot l - 1 = 6 - 1 = 5$.

Ш а г 5. Вычисляем средние квадраты:

$$MS_A = \frac{SS_A}{df_A} = \frac{7{,}5}{1} = 7{,}5;\ MS_B = \frac{SS_B}{df_B} = \frac{5}{2} = 2{,}5;\ MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{45}{2} = 22{,}5;$$

$$MS_{wg} = \frac{SS_{wg}}{df_{wg}} = \frac{52}{24} = 2{,}167;\ MS_{bg} = \frac{SS_{bg}}{df_{bg}} = \frac{57{,}5}{5} = 11{,}5.$$

Ш а г 6. Вычисляем эмпирические значения F-отношения:

$$F_A = \frac{MS_A}{MS_{wg}} = \frac{7{,}5}{2{,}167} = 3{,}46;$$

$$F_B = \frac{MS_B}{MS_{wg}} = \frac{2{,}5}{2{,}167} = 1{,}15;$$

$$F_{AB} = \frac{MS_{AB}}{MS_{wg}} = \frac{22{,}5}{2{,}167} = 10{,}39;$$

$$F_{bg} = \frac{MS_{bg}}{MS_{wg}} = \frac{11{,}5}{2{,}167} = 5{,}31.$$

Ш а г 7. Определяем p-уровень значимости для каждого из F-отношений. Для этого сравниваем эмпирические значения F-отношения с критическими (табличными) для соответствующих чисел степеней свободы по таблице критических значений F-распределения для проверки направленных альтернатив (приложение 3).

$F_A = 3{,}46$; $df_1 = 1$; $df_2 = 24$; $F_{0{,}05} = 4{,}2$; $p > 0{,}05$

$F_B = 1{,}15$; $df_1 = 2$; $df_2 = 24$; $F_{0{,}05} = 3{,}4$; $p > 0{,}05$

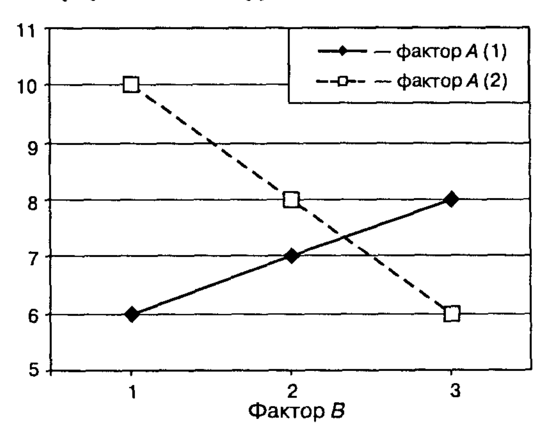
$F_{AB} = 10,39$; $df_1 = 2$; $df_2 = 24$; $F_{0,01} = 5,61$; $p < 0,01$
$F_{bg} = 5,31$; $df_1 = 5$; $df_2 = 24$; $F_{0,01} = 3,90$; $p < 0,01$

Представим результаты в виде таблицы:

Источник изменчивости	Сумма квадратов (*SS*)	*df*	Средний квадрат (*MS*)	*F*	*p*-уровень
Модельная (факторная)	57,5	5	11,5	5,31	<0,01
Фактор *A*	7,5	1	7,5	3,46	>0,05
Фактор *B*	5	2	2,5	1,15	>0,05
A × *B*	45	2	22,5	10,39	<0,01
Ошибка	52	24	2,167	—	—
Общая	109,5	29		—	—

$R^2 = 0,525$

Ш а г 8. Принимаем статистические решения и формулируем содержательные выводы. H_0 на уровне $\alpha = 0,05$ отклоняется в отношении взаимодействия факторов и общего влияния факторов. Обнаружено статистически достоверное совокупное влияние численности группы и наличия (отсутствия) лидера на успешность группового решения задачи ($p < 0,01$). Факторная модель объясняет 52,5% общей доли изменчивости времени решения задачи. Статистически достоверным является взаимодействие фактора лидерства и численности группы ($p < 0,01$). График средних значений позволяет дать интерпретацию обнаруженного взаимодействия:



Чем больше численность группы, тем быстрее решается задача при наличии лидера; без лидера успешнее работают группы меньшей численности.

ANOVA с количеством факторов больше двух принципиально не отличается от двухфакторного варианта. Специфика ANOVA с числом факторов больше двух заключается в наличии проблемы взаимодействия более чем двух факторов. В двухфакторном случае анализируется *взаимодействие первого порядка* (двух факторов). А в трехфакторном ANOVA, с факторами *A*, *B* и *C*, помимо двухфакторных взаимодействий (первого порядка) $A \times B$, $A \times C$ и $B \times C$ необходимо рассматривать и трехфакторное *взаимодействие второго порядка*: $A \times B \times C$.

ПРИМЕР 13.7

Предположим, при изучении влияния численности группы и наличия или отсутствия в ней лидера на успешность решения задачи введен еще один фактор — тип задания (фактор *A* — наличие лидера, две градации: 1 — нет лидера, 2 — есть лидер; фактор *B* — численность группы, три градации: 1 — 2–3 человека, 2 — 5–7 человек, 3 — 10–15 человек; фактор *C* — тип задания, две градации: 1 — групповое задание, 2 — индивидуальное задание). Графики средних значений (рис. 13.3) демонстрируют трехфакторное взаимодействие (второго порядка): взаимодействие факторов *A* и *B* проявляется по-разному в зависимости от градаций фактора *C*. Обратите внимание, что это взаимодействие допускает три эквивалентные формы интерпретации: а) тип задания по-разному влияет на успешность в зависимости от численности группы и наличия или отсутствия лидера; б) численность группы по-разному влияет на успешность решения задачи в зависимости от типа задания и наличия или отсутствия лидера в группе; в) наличие или отсутствие лидера по-разному влияет на успешность решения задачи в зависимости от численности группы и типа задания. Обратите также внимание на то, насколько сложна более детальная интерпретация взаимодействия второго порядка, по сравнению с интерпретацией взаимодействия первого порядка.

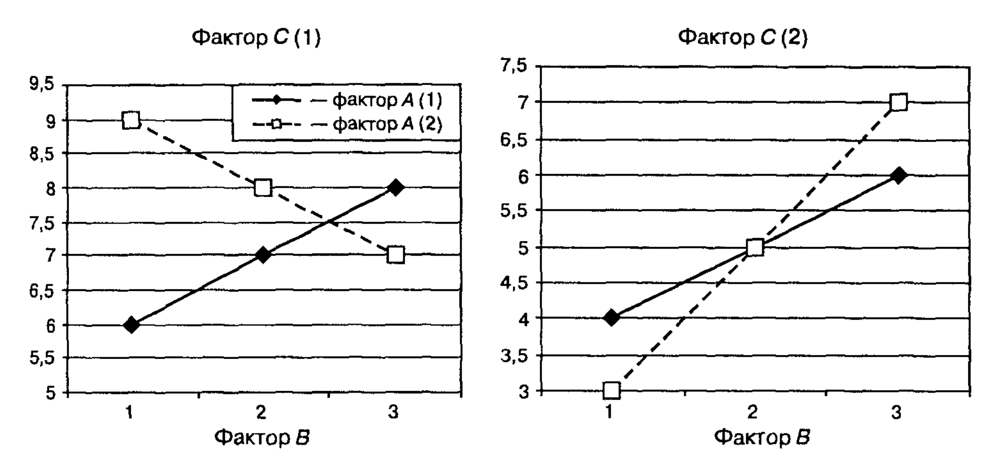


Рис. 13.3. Взаимодействие факторов *A* и *B* на разных уровнях фактора *C*

Пример демонстрирует трудности, связанные с интерпретацией трехфакторного взаимодействия. Интерпретация взаимодействий более высокого порядка еще сложнее, если вообще возможна. Ситуацию осложняет и то, что количество взаимодействий с увеличением числа факторов растет в геометрической прогрессии: количество проверяемых гипотез в ANOVA — о главных эффектах и всех взаимодействиях факторов выражается формулой:

$$K = 2^P - 1,$$

где P — число факторов, K — количество проверяемых гипотез. Так, если двухфакторный ANOVA предполагает проверку трех гипотез, то трехфакторный — уже семи, а четырехфакторный — 15-ти. Поэтому без острой необходимости нежелательно включать в ANOVA более трех факторов.

Обработка на компьютере

Рассмотрим применение однофакторного ANOVA на примере изучения влияния численности группы и наличия или отсутствия лидера в группе на успешность группового решения задачи (данные примера 13.6).

Исходные данные для анализа введены в таблицу (**Data Editor**) в следующем виде:

	time	factor_a	factor_b
1	4	1	1
2	8	1	1
3	5	1	1
...	...	...	...
29	6	2	3
30	8	2	3

Каждой строке соответствует одна группа; `time` — зависимая переменная (время); `factor_a` — наличие лидера (1 — нет, 2 — есть); `factor_b` — размер группы (1 — 2–3 человека, 2 — 5–7 человек, 3 — 10–15 человек).

1. Выбираем **Analyze > General Linear Model > Univariate...**

2. Задаем зависимую переменную и факторы. В открывшемся окне диалога выделяем и переносим при помощи кнопки ▹ из левого окна зависимую переменную в правое верхнее окно (**Dependent Variables**); переменные, соответствующие факторам, — в правое второе сверху окно (**Fixed Factor(s)**).

3. Задаем дополнительные опции: описательные статистики и проверку однородности дисперсии. Нажимаем кнопку **Options...** (Опции) и в открывшемся окне отмечаем флажком **Descriptive Statistics** (Описательные статистики), **Homogeneity tests** (Тесты однородности дисперсии). Нажимаем **Continue** (Продолжить).

4. Задаем вид графиков средних значений. Нажимаем кнопку **Plots** (Графики). В открывшемся окне диалога задаем имя фактора, соответствующего горизонтальной оси графика (того, который имеет больше градаций): выделяем в левом окне **factor_b** и переносим в верхнее правое окно (**Horizontal Axis**) при помощи кнопки ▹. Присваиваем имя второго фактора отдельным линиям на графике: выделяем в левом окне **factor_a** и переносим его во второе сверху правое окно (**Separate Lines**). Нажимаем **Plots: Add** (в нижнем окне появляется **factor_b* factor_a**). Нажимаем **Continue** (Продолжить).

(Как и для однофакторного ANOVA, можно было бы воспользоваться функциями **Post Hoc** (Множественные сравнения) и **Contrasts** (Контрасты), но мы этого сейчас не делаем.) Нажимаем OK.

5. Получаем результаты.

А) Описательные статистики:

Descriptive Statistics
Dependent Variable: TIME

FACTOR_A	FACTOR_B	Mean	Std. Deviation	N
1.00	1.00	6.0000	1.58114	5
	2.00	7.0000	1.58114	5
	3.00	8.0000	1.22474	5
	Total	7.0000	1.60357	15
2.00	1.00	10.0000	1.58114	5
	2.00	8.0000	1.22474	5
	3.00	6.0000	1.58114	5
	Total	8.0000	2.17124	15
Total	1.00	8.0000	2.58199	10
	2.00	7.5000	1.43372	10
	3.00	7.0000	1.69967	10
	Total	7.5000	1.94316	30

B) Тест однородности дисперсии:

Levene's Test of Equality of Error Variances(a)
Dependent Variable: TIME

F	df1	df2	Sig.
.305	5	24	.905

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.
a Design: Intercept+FACTOR_A+FACTOR_B+FACTOR_A * FACTOR_B

По критерию Ливена статистически достоверных различий между дисперсиями не обнаружено. Следовательно, применение ANOVA является корректным.

C) Основные результаты дисперсионного анализа:

Tests of Between-Subjects Effects
Dependent Variable: TIME

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	57.500(a)	5	11.500	5.308	.002
Intercept	1687.500	1	1687.500	778.846	.000
FACTOR_A	7.500	1	7.500	3.462	.075
FACTOR_B	5.000	2	2.500	1.154	.332
FACTOR_A * FACTOR_B	45.000	2	22.500	10.385	.001
Error	52.000	24	2.167		
Total	1797.000	30			
Corrected Total	109.500	29			

a R Squared = .525 (Adjusted R Squared = .426)

Результаты соответствуют тем, которые были получены при расчетах «вручную». Факторная модель (Corrected Model) статистически достоверна и объясняет 52,5% общей дисперсии, о чем свидетельствует коэффициент детерминации (R Squared = .525). Влияние факторов *A* и *B* статистически не подтверждается, но статистически достоверно их взаимодействие (Sig. = 0.001).

D) График средних значений:



Profile Plots

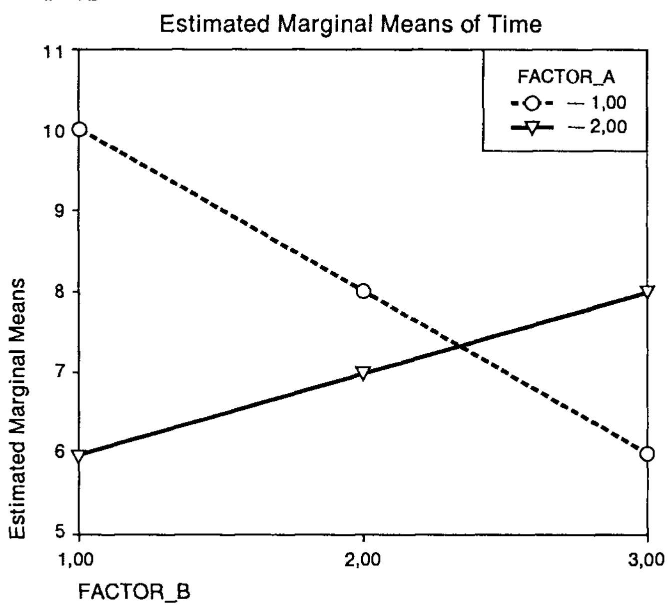


График средних значений существенно облегчает интерпретацию статистически достоверных результатов ANOVA.

ANOVA С ПОВТОРНЫМИ ИЗМЕРЕНИЯМИ

Рассмотренные ранее варианты ANOVA применяются, когда разным градациям изучаемых факторов соответствуют разные группы объектов (испытуемых). Однако часто используются планы исследования, когда разным градациям фактора соответствует одна и та же группа объектов (зависимые выборки). В соответствии с этим различают межгрупповые и внутригрупповые факторы. Разным градациям межгруппового фактора (Between-subject Factor) соответствуют разные группы объектов, а разным градациям внутригруппового фактора (Within-subject Factor) соответствует одна и та же группа объектов (или зависимые выборки).

ANOVA с повторными измерениями (Repeated Measures ANOVA или *GLM Repeated Measures)* применяется, когда по крайней мере один из факторов из-

меняется по *внутригрупповому плану*, то есть разным градациям этого фактора соответствует одна и та же выборка объектов (испытуемых). Конечно, эти выборки можно рассматривать как независимые и применять обычный вариант ANOVA. Но ANOVA с повторными измерениями имеет в этом случае существенное преимущество: он позволяет исключить из общей дисперсии данных *ту ее часть, которая обусловлена индивидуальными различиями в уровне зависимой переменной*. За счет этого метод оказывается более чувствительным к влиянию изучаемых факторов и позволяет с большей надежностью обнаруживать их эффекты.

Таким образом, специфика ANOVA с повторными измерениями заключается в том, что из остаточной изменчивости (внутригрупповой) вычитается компонент, обусловленный индивидуальными различиями. Тем самым уменьшается дисперсия ошибки факторной модели и повышается чувствительность метода к воздействию факторов на зависимую переменную. В остальном, в частности — в отношении проверяемых гипотез, данный вариант ANOVA сохраняет сходство с рассмотренными выше методами ANOVA.

Структура исходных данных: градациям внутригруппового фактора соответствует неоднократное измерение зависимой переменной для одной и той же группы объектов. Допускается наличие межгрупповых факторов, а также нескольких внутригрупповых факторов.

ПРИМЕР

Изучалось влияние интонации на запоминание слов. В качестве материала использовался список из 24 не связанных по смыслу слов одинаковой длины и частоты встречаемости. Одной группе испытуемых весь список читался с неизменной интонацией, а другой — с интонационным выделением серединной восьмерки слов. Зависимой переменной выступало количество правильно воспроизведенных испытуемыми слов: из первых восьми слов ряда, из серединной и из последней восьмерки слов. Предполагалось, что во второй группе будет менее выражен эффект конца и начала ряда, то есть лучше запомнится интонационно выделенная середина ряда. Таким образом, план эксперимента включал 2 фактора: фактор A (внутригрупповой) — часть ряда (три градации); фактор B (межгрупповой) — интонационное выделение (две градации).

Таблица исходных данных:

Испытуемый №	Фактор A (часть ряда)			Фактор B (группа)
	начало	середина	конец	
1	4	5	6	2
2	6	3	5	1
3	3	6	5	2
...	...	...	...	...
N	5	5	7	1

Так же, как и в случае двух межгрупповых факторов, ANOVA с одним межгрупповым и одним внутригрупповым факторами позволяет проверить три

гипотезы: а) эффект внутригруппового фактора *A*; б) эффект межгруппового фактора *B*; в) эффект взаимодействия факторов $A \times B$.

Исходные предположения и, соответственно, *ограничения* на применение ANOVA с повторными измерениями зависят от того, какая из двух моделей используется: одномерная или многомерная. *Одномерная модель* основана на предположении, что каждому уровню внутригруппового фактора соответствует повторное измерение одной и той же зависимой переменной (следовательно, эти измерения положительно коррелируют). *Многомерная модель* свободна от допущения о коррелированности измерений зависимой переменной. Общим для той и другой модели является исходное допущение о том, что множество измерений зависимой переменной для каждого испытуемого является выборкой из многомерного нормального распределения.

Одномерный подход (Univariate approach) основан на применении *F*-отношения, свойственного и другим методам ANOVA. Однако его применение ограничено так называемым допущением о *сферичности ковариационно-дисперсионной матрицы*. Это допущение подразумевает, во-первых, что дисперсии зависимой переменной для разных уровней внутригруппового фактора не отличаются; во-вторых, корреляции между повторными измерениями есть, и они положительны. Для проверки этого предположения в компьютерных программах используется *тест сферичности ковариационно-дисперсионной матрицы Моучли (Mauchly's Test of Sphericity)*. Если тест Моучли показывает статистически достоверный результат, то предположение о сферичности считается ошибочным, и одномерный подход неприменим. Однако на небольших выборках тест сферичности Моучли имет малую чувствительность, а для больших выборок даже небольшие отклонения от сферичности дают статистически значимые результаты. При нарушении допущения о сферичности компьютерные программы предлагают специальную поправку (*эпсилон-коррекцию, Epsilon Corrected*) числа степеней свободы и, соответственно, уровня значимости.

Если предположение о сферичности не отклоняется (результат теста Моучли статистически не достоверен), то более предпочтительным является одномерный подход, как более чувствительный к действию внутригруппового фактора. Если предположение о сферичности отклоняется (результат теста Моучли статистически достоверен), то можно воспользоваться поправками, предлагаемыми компьютерной программой (эпсилон-коррекция). Но более корректно применить многомерный подход.

Многомерный подход (Multivariate approach) свободен от предположения о сферичности, свойственного одномерному подходу. В этом случае используется не *F*-критерий, а *многомерные тесты*, наиболее распространенные из которых *«След Пиллая» (Pillai's Trace)* и *«λ-Вилкса»*[1] *(Wilks' Lambda)*. При использовании межгрупповых факторов дополнительно проверяется допущение об идентичности ковариационно-дисперсионных матриц, соответствующих разным уровням межгрупповых факторов. Это допущение аналогично требованию однородности дисперсии в ANOVA с межгрупповыми факторами, но для

[1] Читается как «лямбда Вилкса».

его проверки в ANOVA с повторными измерениями обычно используется *M-тест Бокса (Box's M-test)*. Если *M*-тест Бокса показывает статистически значимый результат, то дисперсионно-ковариационные матрицы не идентичны, и применение многомерного подхода в этом случае не корректно.

Последовательность ANOVA с повторными измерениями рассмотрим сначала на примере с одним внутригрупповым фактором. Общая изменчивость зависимой переменной (SS_{total}) в этом случае раскладывается на три составляющие:

$$SS_{total} = SS_F + SS_I + SS_{er},$$

где SS_F — факторная изменчивость (между уровнями); SS_I — межиндивидуальная изменчивость (между средними для каждого объекта — испытуемого); SS_{er} — остаточная изменчивость (ошибка).

ПРИМЕР 13.8

Предположим, изучается эффективность воспроизведения предъявленного ряда из 24 не связанных по смыслу слов. Исследователя интересует, будет ли в этом случае проявляться эффект начала и конца ряда. Соответственно, для каждого испытуемого подсчитывалась частота воспроизведения слов из первой, второй и третьей части ряда. Всего в эксперименте участвовало 5 человек. Исходные данные представлены в таблице:

Испытуемый №	Фактор *A* (часть ряда)			Средние
	начало (1)	середина (2)	конец (3)	
1	4	3	6	$M_1 = 4{,}33$
2	6	4	5	$M_2 = 5{,}00$
3	5	5	7	$M_3 = 5{,}67$
4	7	5	7	$M_4 = 6{,}33$
5	3	3	5	$M_5 = 3{,}67$
Средние	$M_{A1} = 5$	$M_{A2} = 4$	$M_{A3} = 6$	$M = 5$

Число объектов (испытуемых): $N = 5$.

Число градаций внутригруппового фактора *A*: $k = 3$.

Шаг 1. Подсчитываем общую сумму квадратов.

$$SS_{total} = \sum_{i}^{k,N}(x_i - M)^2 = (4-5)^2 + (3-5)^2 + (6-5)^2 + (6-5)^2 + (4-5)^2 + (5-5)^2 + \ldots = 28.$$

Шаг 2. Подсчитываем факторную сумму квадратов — между уровнями.

$$SS_F = N\sum_{j}^{k}(x_j - M)^2 = 5\cdot\left[(5-5)^2 + (4-5)^2 + (6-5)^2\right] = 10.$$

Шаг 3. Подсчитываем межиндивидуальную сумму квадратов.

$$SS_I = k\sum_{i}^{N}(x_i - M)^2 = 3\cdot\left[(4{,}33-5)^2 + (5-5)^2 + (5{,}67-5)^2 + (6{,}33-5)^2 + (3{,}67-5)^2\right] = 13{,}33.$$

Шаг 4. Подсчитываем остаточную сумму квадратов.

$$SS_{er} = SS_{total} - SS_F - SS_I = 28 - 10 - 13{,}33 = 4{,}67.$$

Шаг 5. Определяем числа степеней свободы для сумм квадратов:
для общей: $df_{total} = N \cdot k - 1 = 5 \cdot 3 - 1 = 14$;
для фактора: $df_F = k - 1 = 3 - 1 = 2$;
для остаточной: $df_{er} = (N - 1)(k - 1) = 8$.

Шаг 6. Вычисляем средние квадраты.

$$MS_F = \frac{SS_F}{df_F} = \frac{10}{2} = 5;\; MS_{er} = \frac{SS_B}{df_B} = \frac{4{,}67}{8} = 0{,}583.$$

Шаг 7. Вычисляем эмпирическое значение F-отношения:

$$F_F = \frac{MS_F}{MS_{er}} = \frac{5}{0{,}583} = 8{,}571.$$

Шаг 8. Определяем p-уровень значимости для F-отношения. Для этого сравниваем эмпирическое значение F-отношения с критическими (табличными) для соответствующих чисел степеней свободы по таблице критических значений F-распределения для проверки направленных альтернатив (приложение 3).

$$F_F = 8{,}571;\; df_1 = 2;\; df_2 = 8;\; F_{0{,}05} = 4{,}46;\; p < 0{,}05.$$

Представим результаты в виде таблицы:

Источник изменчивости	Сумма квадратов (SS)	df	Средний квадрат (MS)	F	p-уровень
Факторный	10	2	5	8,571	<0,05
Ошибки	4,67	8	0,583	—	—
Общий	28	14	—	—	—
Межиндивидуальный	13,33	—	—	—	—

Шаг 9. Принимаем статистические решения и формулируем содержательные выводы. H_0 на уровне $\alpha = 0{,}05$ отклоняется. Обнаружено статистически достоверное влияние фактора положения слова в ряду на его запоминание ($p < 0{,}05$).

Заметим, что если в последнем примере рассматривать повторные измерения как независимые группы и провести однофакторный ANOVA (с межгрупповым фактором A), то статистически значимое влияние фактора *обнаружено не будет* — индивидуальные различия между испытуемыми «перекроют» факторный эффект.

Двухфакторный ANOVA с повторными измерениями по одному из факторов (с одним внутригрупповым и одним межгрупповым факторами) позволяет проверить три гипотезы: а) о влиянии внутригруппового фактора; б) о влиянии межгруппового фактора; в) о взаимодействии внутригруппового и межгруппового факторов (о зависимости влияния межгруппового фактора от уровней внутригруппового фактора — или наоборот).

Этот вариант ANOVA имеет свою специфику, связанную с выделением составных частей общей изменчивости зависимой переменной. Рассмотрим соотношение различных источников изменчивости на примере исследования влияния интонации на запоминание ряда слов.

ПРИМЕР 13.9

В эксперименте участвовало 2 группы испытуемых (фактор *A* — межгрупповой, два уровня): 1 — все 24 слова ряда предъявлялись с одинаковой интонацией; 2 — серединная восьмерка из того же предъявляемого ряда слов интонационно выделялась. Для каждого испытуемого измерялось по три показателя зависимой переменной — количества воспроизведенных слов (фактор *B* — внутригрупповой, три градации): из первой, второй и третьей восьмерки предъявленных слов.

Результаты эксперимента представлены в таблице:

Испытуемый	Фактор *A* (Уровень 1)			Средние	Испытуемый	Фактор *A* (Уровень 2)			Средние
	1	2	3			1	2	3	
1	4	3	6	4,33	6	3	5	4	4,00
2	6	4	5	5,00	7	5	6	3	4,67
3	5	5	7	5,67	8	4	7	5	5,33
4	7	5	7	6,33	9	6	7	5	6,00
5	3	3	5	3,67	10	2	5	3	3,33
Средние:	5	4	6	$M_{A1} = 5$	Средние:	4	6	4	$M_{A2} = 4{,}67$
$M_{B1} = 4{,}5$; $M_{B2} = 5$; $M_{B3} = 5$									
$M = 4{,}833$; $\sigma^2 = 2{,}075$									

Модель двухфакторного ANOVA с межгрупповым и внутригрупповым факторами предполагает разделение общей изменчивости данных на две составляющие: а) изменчивость между объектами или межиндивидуальная изменчивость (SS_{bs}); б) внутригрупповая изменчивость (SS_{wg}).

$$SS_{total} = SS_{bs} + SS_{wg}.$$

Межиндивидуальная изменчивость состоит из изменчивости между градациями межгруппового фактора (SS_A) и изменчивости между испытуемыми внутри этих градаций (SS_{IWG}), или, что то же самое, из изменчивости средних значений для каждого испытуемого относительно общего среднего.

$$SS_{bs} = SS_A + SS_{IWG} = n \cdot l \cdot \sum_{j=1}^{k}(M_j - M)^2 + l \cdot \sum_{j=1}^{k}\sum_{i=1}^{n}(M_i - M_k)^2,$$

или

$$SS_{bs} = l \cdot \sum_{i=1}^{n \cdot k}(M_i - M)^2,$$

где n — численность объектов в одной градации межгруппового фактора; k — число градаций межгруппового фактора; l — число градаций внутригруппового фактора. SS_{IWG} — это мера ошибки межгрупповой факторной модели, или фактора *B*.

Внутригрупповая изменчивость — это сумма трех составляющих изменчивости: а) под влиянием внутригруппового фактора *B* (SS_B); б) под влиянием взаимодействия межгруппового и внутригруппового факторов (SS_{AB}); в) остаточной внутригрупповой изменчивости — ошибки модели (SS_{erB}).

$$SS_{wg} = SS_B + SS_{AB} + SS_{erB}.$$

SS_B вычисляется как сумма квадратов, обусловленная различиями l средних значений (для градаций внутригруппового фактора) относительно общего среднего значения. SS_{AB} вычисляется как сумма квадратов, обусловленная различиями $k \times l$ средних относительно общего среднего.

ПРИМЕР 13.9 (продолжение)

Шаг 1. Вычисляем SS_{total}, SS_A, SS_{bs} и SS_{wg}:

$$SS_{total} = (N-1) \cdot \sigma^2 = 29 \cdot 2,075 = 60,175;$$

$$SS_A = n \cdot l \cdot \sum_{i=1}^{k}(M_i - M)^2 = 5 \cdot 3 \cdot [(5-4,83)^2 + (4,67-4,83)^2] = 0,833;$$

$$SS_{IWG} = l \cdot \sum_{j=1}^{k}\sum_{i=1}^{n}(M_i - M_k)^2 = 26,667;$$

$$SS_{bs} = l \cdot \sum_{i=1}^{n \cdot k}(M_i - M)^2 = 3 \cdot \left[(4,33-4,833)^2 + (5-4,833)^2 + \ldots + (3,33-4,833)^2\right] = 27,5;$$

$$SS_{wg} = SS_{total} - SS_{bs} = 60,175 - 27,5 = 32,675.$$

Шаг 2. Вычисляем SS_B и SS_{AB}:

$$SS_B = N \cdot \sum_{i=1}^{l}(M_l - M)^2 = 10 \cdot \left[(4,5-4,833)^2 + (5-4,833)^2 + (5-4,833)^2\right] = 1,667;$$

$$SS_{AB} = n \cdot \sum_{i=1}^{k}\sum_{j=1}^{l}(M_{ij} - M)^2 - SS_A - SS_B =$$

$$= 5 \cdot [(5-4,83)^2 + (4-4,83)^2 + (6-4,83)^2 + \ldots + (4-4,83)^2] - 0,83 - 1,67 = 21,67.$$

Шаг 3. Вычисляем остаточную сумму квадратов SS_{erB}:

$$SS_{erB} = SS_{wg} - SS_B - SS_{AB} = 32,675 - 1,667 - 21,67 = 9,33.$$

Шаг 4. Определяем числа степеней свободы:

$$df_A = k - 1 = 1;$$
$$df_B = l - 1 = 2;$$
$$df_{IWG} = N - k = 8;$$
$$df_{AB} = (k-1)(l-1) = 2;$$
$$df_{erB} = (N-k)(l-1) = 16.$$

Шаг 5. Вычисляем средние квадраты:

$$MS_A = \frac{SS_A}{df_A} = \frac{0,833}{1} = 0,833;$$

$$MS_{IWG} = \frac{SS_{IWG}}{df_{IWG}} = \frac{26,67}{8} = 3,33;$$

$$MS_B = \frac{SS_B}{df_B} = \frac{1,67}{2} = 0,833;$$

$$MS_{erB} = \frac{SS_{erB}}{df_{erB}} = \frac{9{,}33}{16} = 0{,}58\ ;$$

$$MS_{AB} = \frac{SS_{AB}}{df_{AB}} = \frac{21{,}67}{2} = 10{,}84\ .$$

Ш а г 6. Вычислим F-отношения.

$$F_A = \frac{MS_A}{MS_{IWG}} = \frac{0{,}833}{3{,}33} = 0{,}25\ ;$$

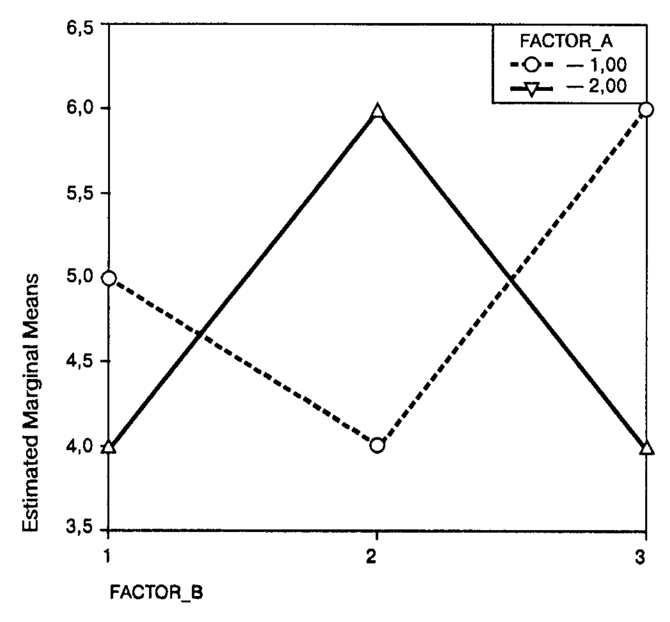
$$F_B = \frac{MS_B}{MS_{erB}} = \frac{0{,}833}{0{,}58} = 1{,}43\ ;$$

$$F_{AB} = \frac{MS_{AB}}{MS_{erB}} = \frac{10{,}84}{0{,}58} = 18{,}57\ .$$

Ш а г 7. Определяем уровень значимости и представляем результаты в виде таблицы:

Источник изменчивости	Сумма квадратов (SS)	df	Средний квадрат (MS)	F	p-уровень
Фактор A	0,833	1	0,833	0,25	>0,05
Фактор B	1,67	2	0,833	1,43	>0,05
$A \times B$	21,67	2	10,84	18,57	<0,01
Ошибка межгрупповая	26,67	8	3,33	—	—
Ошибка внутригрупповая	9,33	16	0,58	—	—

Ш а г 8. Принимаем статистические решения и формулируем выводы. H_0 отклоняется только в отношении взаимодействия факторов. Обнаружено взаимодействие



интонационного выделения и положения слова в ряду ($p < 0,01$): влияние интонационного выделения середины ряда на эффективность воспроизведения слов зависит от того, в какой части ряда находятся слова. График средних значений позволяет дать более детальную интерпретацию взаимодействия: середина ряда при интонационном выделении запоминается лучше краев ряда, а без интонационного выделения — наоборот: слова в начале и конце ряда запоминаются лучше, чем в середине ряда.

Обработка на компьютере

Используем для компьютерной обработки те же исходные данные, по которым вычисления ранее производились «вручную» (пример 13.9).

Исходные данные для анализа введены в таблицу (**Data Editor**) в следующем виде:

№	v1	v2	v3	factor_a
1	4	3	6	1
2	6	4	5	1
3	5	5	7	1
4	7	5	7	1
5	3	3	5	1
6	3	5	4	2
7	5	6	3	2
8	4	7	5	2
9	6	7	5	2
10	2	5	3	2

Зависимая переменная — количество правильно воспроизведенных слов каждым из 10 испытуемых. Фактор *A* — условия предъявления (межгрупповой), имеет 2 градации и ему соответствует номинативная переменная factor_a. Фактор *B* — часть ряда (внутригрупповой), имеет 3 градации, которым соответствуют 3 переменные: v1 — начало ряда, v2 — середина ряда, v3 — конец ряда.

1. Выбираем **Analyze > General Linear Model > Repeated Measures...**

2. В открывшемся окне диалога **Repeated Measures Define Factor(s)** (Определение факторов для повторных измерений): а) задаем имя фактора в окне **Within-Subjects Factor Name** (Имя внутригруппового фактора): factor_b; б) задаем число уровней этого фактора в окне **Number of Levels** (Число уровней): 3. Нажимаем **Add** (Добавить). В принципе, возможно добавление еще и других внутригрупповых факторов, но в нашем случае в этом нет необходимости. Поэтому нажимаем **Define** (Определить).

3. В открывшемся окне диалога **Repeated Measures** (Повторные измерения) задаем уровни внутригруппового фактора и межгрупповой фактор. Выделя-

ем и переносим из левого окна при помощи кнопки ▷: переменные v1, v2, v3 — в правое верхнее окно **Within-Subjects Variables** (Внутригрупповые переменные); переменную factor_a — во второе правое окно **Between-Subjects Factor**.

4. Задаем необходимые опции в окне диалога **Repeated Measures** (Повторные измерения). Нажимаем **Options...** (Опции). В открывшемся окне отмечаем флажком **Descriptive Statistics** (Описательные статистики) и **Homogeneity Tests** (Тесты однородности дисперсии). Нажимаем **Continue** (Продолжить).

5. Задаем вид графиков средних значений в окне диалога **Repeated Measures** (Повторные измерения). Нажимаем **Plots...** (Графики). В открывшемся окне диалога задаем имя фактора, соответствующего горизонтальной оси графика (того, который имеет больше градаций): выделяем в левом окне factor_b и переносим в верхнее правое окно (**Horizontal Axis**) при помощи кнопки ▷. Присваиваем имя второго фактора отдельным линиям на графике: выделяем в левом окне factor_a и переносим его во второе сверху правое окно (**Separate Lines**). Нажимаем **Plots: Add** (в нижнем окне появляется factor_b* factor_a). Нажимаем **Continue** (Продолжить). (Как и для других вариантов ANOVA, можно было бы воспользоваться функциями **Post Hoc** (Множественные сравнения) и **Contrasts** (Контрасты) для межгруппового фактора — если имеется более двух градаций, но мы этого сейчас не делаем.) Нажимаем OK.

6. Получаем результаты.

А) Описательные статистики:

Descriptive Statistics

	FACTOR_A	Mean	Std. Deviation	N
V1	1.00	5.0000	1.58114	5
	2.00	4.0000	1.58114	5
	Total	4.5000	1.58114	10
V2	1.00	4.0000	1.00000	5
	2.00	6.0000	1.00000	5
	Total	5.0000	1.41421	10
V3	1.00	6.0000	1.00000	5
	2.00	4.0000	1.00000	5
	Total	5.0000	1.41421	10

В) Результаты М-теста Бокса (**Box's M-test**).

Box's Test of Equality of Covariance Matrices(a)

Box's M	.000
F	.000
df1	6
df2	463.698
Sig.	1.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a Design: Intercept+FACTOR_A Within Subjects Design: FACTOR_B

Тест показывает, что дисперсионно-ковариационные матрицы, соответствующие разным градациям межгруппового фактора, статистически достоверно не отличаются друг от друга. Следовательно, применение многомерного подхода является корректным.

C) Результаты многомерных тестов.

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
FACTOR_B	Pillai's Trace	.263	1.250(a)	2.000	7.000	.343
	Wilks' Lambda	.737	1.250(a)	2.000	7.000	.343
	Hotelling's Trace	.357	1.250(a)	2.000	7.000	.343
	Roy's Largest Root	.357	1.250(a)	2.000	7.000	.343
FACTOR_B* FACTOR_A	Pillai's Trace	.950	66.250(a)	2.000	7.000	.000
	Wilks' Lambda	.050	66.250(a)	2.000	7.000	.000
	Hotelling's Trace	18.929	66.250(a)	2.000	7.000	.000
	Roy's Largest Root	18.929	66.250(a)	2.000	7.000	.000

a Exact statistic

b Design: Intercept+FACTOR_A Within Subjects Design: FACTOR_B

Многомерные тесты не показывают статистически достоверного влияния внутригруппового фактора *B*. Но взаимодействие факторов *A* и *B* оказывается достоверным на высоком уровне статистической значимости.

D) Результаты теста сферичности Моучли:

Mauchly's Test of Sphericity(b)

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon(a)		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
FACTOR_B	.429	5.931	2	.052	.636	.797	.500

a May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b Design: Intercept+FACTOR_A Within Subjects Design: FACTOR_B

Результат теста сферичности Моучли не достигает статистически значимого уровня (`Sig. > 0.05`). Следовательно, дисперсии для уровней внутригруппового фактора на разных уровнях межгруппового фактора существенно не отличаются, корреляции между повторными измерениями есть, но они не достигают единицы. Применение одномерного подхода является корректным.

E) Результаты одномерных тестов эффекта внутригруппового фактора *B*:

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
FACTOR_B	Sphericity Assumed	1.667	2	.833	1.429	.269
	Greenhouse-Geisser	1.667	1.273	1.310	1.429	.270
	Huynh-Feldt	1.667	1.595	1.045	1.429	.270
	Lower-bound	1.667	1.000	1.667	1.429	.266
FACTOR_B * FACTOR_A	Sphericity Assumed	21.667	2	10.833	18.571	.000
	Greenhouse-Geisser	21.667	1.273	17.024	18.571	.001
	Huynh-Feldt	21.667	1.595	13.588	18.571	.000
	Lower-bound	21.667	1.000	21.667	18.571	.003
Error (FACTOR_B)	Sphericity Assumed	9.333	16	.583		
	Greenhouse-Geisser	9.333	10.182	.917		
	Huynh-Feldt	9.333	12.757	.732		
	Lower-bound	9.333	8.000	1.167		

Это результаты проверки гипотез относительно внутригрупповых эффектов факторов с применением *F*-отношения. Они идентичны полученным при вычислениях «вручную» (пример 13.9, шаг 6) для фактора *B* и взаимодействия *A* и *B*. Эффект фактора *B* статистически не достоверен, но взаимодействие факторов *A* и *B* статистически значимо.

F) Результаты проверки однородности дисперсии по критерию Ливена:

Levene's Test of Equality of Error Variances(a)

	F	df1	df2	Sig.
V1	.000	1	8	1.000
V2	.000	1	8	1.000
V3	.000	1	8	1.000

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a Design: Intercept+FACTOR_A Within Subjects Design: FACTOR_B

Критерий Ливена применяется в данном случае для сравнения градаций межгруппового фактора в отношении каждого повторного измерения зависимой переменной. Результаты демонстрируют отсутствие статистически достоверных различий дисперсий для каждого из трех измерений зависимой переменной. Следовательно, эффект межгруппового фактора может быть принят во внимание.

G) Результаты одномерного теста эффекта межгруппового фактора *A*:

Tests of Between-Subjects Effects
Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	700.833	1	700.833	210.250	.000
FACTOR_A	.833	1	.833	.250	.631
Error	26.667	8	3.333		

Это результат проверки гипотезы о влиянии межгруппового фактора *A* с применением *F*-отношения. Он также идентичен полученным при вычислениях «вручную» (шаг 6) для фактора *A*. Эффект фактора *A* статистически не достоверен.

Н) График средних значений:

Profile Plots

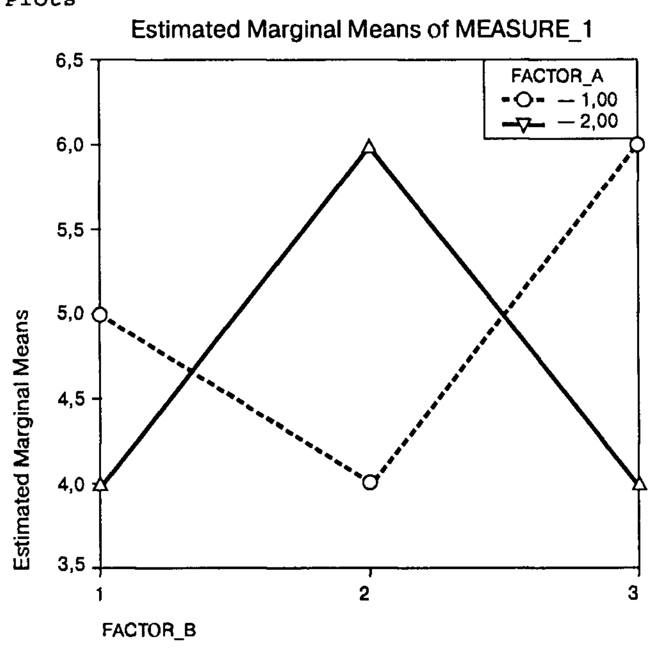


График средних значений облегчает интерпретацию полученных результатов. Середина ряда при интонационном выделении запоминается лучше краев ряда, а без интонационного выделения — наоборот: слова в начале и в конце ряда запоминаются лучше, чем в середине ряда.

МНОГОМЕРНЫЙ ANOVA (MANOVA)

Многомерный ANOVA применяется для изучения эффектов влияния факторов не на одну, а на несколько зависимых переменных (на *многомерную зависимую переменную*). Таким образом, для каждого объекта (испытуемого) имеют-

ся несколько зависимых переменных, которые подвергаются дисперсионному анализу. Поскольку зависимых переменных несколько, то общепринятое его сокращенное обозначение — *MANOVA (Multivariate ANOVA)*. MANOVA позволяет проверить не только гипотезы о влиянии факторов на каждую из зависимых переменных в отдельности, но и гипотезу о влиянии факторов на всю совокупность зависимых переменных, как на одну многомерную переменную.

Структура исходных данных для MANOVA похожа на структуру исходных данных для ANOVA с повторными измерениями. Однако в отличие от ANOVA с повторными измерениями в MANOVA зависимые переменные не обязательно являются повторными измерениями одной и той же переменной, но могут быть и разными переменными. При этом предполагается, что зависимые переменные — это различные измерения одного и того же свойства (явления).

MANOVA может применяться как альтернатива ANOVA с повторными измерениями в случае, если не выполняется его основное допущение — о сферичности ковариационно-дисперсионной матрицы. Однако при выборе MANOVA вместо ANOVA с повторными измерениями необходимо учитывать, что MANOVA является более сложной, но менее мощной (чувствительной) процедурой, особенно в отношении выборок небольшой численности.

Последовательность MANOVA включает в себя два этапа: многомерный и одномерный. *Многомерный подход* применяется для проверки гипотез о влиянии факторов на многомерную зависимую переменную. При этом предполагается, что множество зависимых переменных — это множество измерений одной, но многомерной зависимой переменной. Соответственно, при проверке гипотезы о влиянии факторов на многомерную зависимую переменную учитываются корреляции между различными измерениями этой зависимой переменной. *На одномерном этапе* проверяются гипотезы о влиянии факторов на каждую из зависимых переменных в отдельности. Таким образом, одномерный этап — это реализация обычного ANOVA к каждой из зависимых переменных. Назначение одномерного этапа — детализация результатов многомерного анализа.

Математические допущения MANOVA связаны с тем, что зависимая переменная рассматривается как многомерная величина. Первое допущение — о многомерном нормальном распределении зависимых переменных. Второе допущение — о равенстве дисперсионно-ковариационных матриц для каждого уровня факторов и их сочетаний. Первое допущение не проверяется, так как MANOVA так же устойчив к отклонениям выборочных распределений от нормального вида, как и другие виды ANOVA. Второе допущение эквивалентно допущениям об однородности дисперсии для обычного ANOVA. В данном случае, как и в ANOVA с повторными измерениями, это требование идентичности ковариационно-дисперсионных матриц, соответствующих разным уровням межгрупповых факторов. Для проверки этого допущения так же применяется М-тест Бокса (Box's Test). Дополнительно для одномерного этапа необходимо выполнение допущения об однородности дисперсий, которое проверяется при помощи критерия Ливена (Levene's Test).

Дополнительным условием проведения MANOVA может являться зависимость друг от друга самих зависимых переменных (их корреляция). Для проверки этого условия надо убедиться, что недиагональные элементы корреляционной матрицы зависимых переменных существенно отличаются от нуля. Для статистической проверки допущения о коррелированности зависимых переменных применяется *тест сферичности остатков ковариационной матрицы Бартлетта (Bartlett's Test of Sphericity)*, который проводится, если воспользоваться опцией Residual SSCP matrix (Остатки дисперсионно-ковариационной матрицы) программы SPSS.

Основные показатели MANOVA включают в себя многомерные и одномерные критерии. В качестве многомерных критериев используются *многомерные тесты (Multivariate Tests)*, учитывающие корреляцию зависимых переменных. Обычно вычисляются несколько многомерных критериев, обладающих разной мощностью (чувствительностью). Программа SPSS вычисляет следующие многомерные критерии (в порядке убывания их мощности): Пиллая (Pillai's Trace), Вилкса (Wilks' Lambda), Хотеллинга (Hotelling's Trace) и Роя (Roy's Largest Root). Эти критерии, а также уровни их статистической значимости вычисляются для каждого фактора и всех взаимодействий. *Одномерные критерии (Tests of Between-Subjects Effects)* — это обычные F-отношения для проверки гипотез о влиянии факторов и их взаимодействий на каждую из зависимых переменных в отдельности. Схема проведения MANOVA предполагает, что *одномерные критерии позволяют детализировать те эффекты, статистическая значимость которых подтверждена многомерными критериями.*

Последовательность и основные показатели MANOVA рассмотрим на примере обработки данных гипотетического эксперимента при помощи программы SPSS.

ПРИМЕР 13.10

Предположим, изучалось влияние интонации на запоминание ряда из 24 несвязанных по смыслу слов. Эксперимент состоял из двух серий, в каждой из которых участвовало по 10 испытуемых. В первой серии использовались слова с одинаково высокой, а во второй — с одинаково низкой частотой встречаемости. В каждой серии половине из 10 испытуемых весь ряд предъявлялся с одинаковой интонацией, а половине — с интонационным выделением серединных восьми слов. Затем для каждого испытуемого подсчитывались три показателя количества правильно воспроизведенных слов: из первой, второй и третьей части предъявленного ряда (по 8 слов). Таким образом, эксперимент включал в себя два фактора: фактор A — интонационное выделение (2 градации: 1 — нет, 2 — есть), фактор B — частота встречаемости слов (две градации: 1 — высокочастотные, 2 — низкочастотные). Изучалось влияние этих факторов на три зависимые переменные — показатели успешности воспроизведения слов: $v1$ — для начала ряда, $v2$ — для середины ряда, $v3$ — для конца ряда.

Обработка на компьютере

Исходные данные (пример 13.10) для анализа введены в таблицу (**Data Editor**) в следующем виде:

№	v1	v2	v3	a	b
1	4	3	6	1	1
2	6	4	5	1	1
3	5	5	7	1	1
4	7	5	7	1	1
5	3	3	5	1	1
6	3	5	4	2	1
7	5	6	3	2	1
8	4	7	5	2	1
9	6	7	5	2	1
10	2	5	3	2	1
11	3	2	5	1	2
12	4	3	4	1	2
13	4	4	5	1	2
14	4	4	4	1	2
15	2	2	4	1	2
16	2	4	3	2	2
17	4	5	2	2	2
18	3	4	3	2	2
19	3	5	4	2	2
20	1	4	2	2	2

1. Выбираем **Analyze > General Linear Model > Multivariate...**

2. В открывшемся окне диалога выделяем и переносим при помощи кнопки ▷ из левого окна: зависимые переменные (v1, v2, v3) в правое верхнее окно (**Dependent Variables**); переменные, соответствующие факторам (**a**, **b**) — в правое второе сверху окно (**Fixed Factor(s)**). Нажимаем кнопку **Options...** (Опции) и в открывшемся окне отмечаем флажком **Descriptive Statistics** (Описательные статистики), **Homogeneity tests** (Тесты однородности дисперсии), **Residual SSCP matrix** (Остатки дисперсионно-ковариационной матрицы — для вычисления критерия сферичности Бартлетта.) Нажимаем **Continue** (Продолжить). (Можно также, как и в других методах ANOVA, воспользоваться опциями **Plots** (Графики), **Post Hoc** (Множественные сравнения), **Contrasts** (Контрасты), но мы этого сейчас не делаем). Нажимаем OK.

3. Получаем результаты. Рассмотрим наиболее важные из них.

А) Детальные описательные статистики:

Descriptive Statistics

	A	B	Mean	Std. Deviation	N
V1	1.00	1.00	5.0000	1.58114	5
		2.00	3.4000	.89443	5
		Total	4.2000	1.47573	10
	2.00	1.00	4.0000	1.58114	5
		2.00	2.6000	1.14018	5
		Total	3.3000	1.49443	10

	A	B	Mean	Std. Deviation	N
	Total	1.00	4.5000	1.58114	10
		2.00	3.0000	1.05409	10
		Total	3.7500	1.51744	20
V2	1.00	1.00	4.0000	1.00000	5
		2.00	3.0000	1.00000	5
		Total	3.5000	1.08012	10
	2.00	1.00	6.0000	1.00000	5
		2.00	4.4000	.54772	5
		Total	5.2000	1.13529	10
	Total	1.00	5.0000	1.41421	10
		2.00	3.7000	1.05935	10
		Total	4.3500	1.38697	20
V3	1.00	1.00	6.0000	1.00000	5
		2.00	4.4000	.54772	5
		Total	5.2000	1.13529	10
	2.00	1.00	4.0000	1.00000	5
		2.00	2.8000	.83666	5
		Total	3.4000	1.07497	10
	Total	1.00	5.0000	1.41421	10
		2.00	3.6000	1.07497	10
		Total	4.3000	1.41793	20

B) Результаты М-теста Бокса:

Box's Test of Equality of Covariance Matrices(a)

Box's M	9.520
F	.339
df1	18
df2	904.638
Sig.	.996

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.
a Design: Intercept+A+B+A * B

М-тест Бокса не достигает уровня статистической значимости (Sig. > 0,05), следовательно, дисперсионно-ковариационные матрицы статистически достоверно не различаются. Это значит, что выполняется главное допущение для многомерных тестов и их результаты могут быть приняты к рассмотрению.

C) Результаты теста сферичности Бартлетта:

Bartlett's Test of Sphericity(a)

Likelihood Ratio	.000
Approx. Chi-Square	20.841
df	5
Sig.	.001

Tests the null hypothesis that the residual covariance matrix is proportional to an identity matrix.
a Design: Intercept+A+B+A * B

Тест показывает статистически достоверный результат (Sig. = 0,001). Следовательно, зависимые переменные связаны друг с другом корреляциями.

D) Результаты многомерных тестов:

Multivariate Tests(b)

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.976	192.6(a)	3	14	.000
	Wilks' Lambda	.024	192.6(a)	3	14	.000
	Hotelling's Trace	41.27	192.6(a)	3	14	.000
	Roy's Largest Root	41.27	192.6(a)	3	14	.000
A	Pillai's Trace	.896	40.01(a)	3	14	.000
	Wilks' Lambda	.104	40.01(a)	3	14	.000
	Hotelling's Trace	8.57	40.01(a)	3	14	.000
	Roy's Largest Root	8.57	40.01(a)	3	14	.000
B	Pillai's Trace	.494	4.552(a)	3	14	.020
	Wilks' Lambda	.506	4.552(a)	3	14	.020
	Hotelling's Trace	.975	4.552(a)	3	14	.020
	Roy's Largest Root	.975	4.552(a)	3	14	.020
A * B	Pillai's Trace	.158	.875(a)	3	14	.478
	Wilks' Lambda	.842	.875(a)	3	14	.478
	Hotelling's Trace	.187	.875(a)	3	14	.478
	Roy's Largest Root	.187	.875(a)	3	14	.478

a Exact statistic
b Design: Intercept+A+B+A * B

Многомерные тесты показывают статистически значимые результаты влияния факторов *A* и *B* (Sig. < 0,05); взаимодействие факторов не достигает уровня статистической значимости (Sig. > 0,05). Следовательно, интонационное выделение и частота встречаемости слов статистически достоверно влияет на продуктивность их воспроизведения. Не обнаружена зависимость влияния интонационного выделения от частоты встречаемости слов. Иными словами, интонационное выделение влияет на продуктивность воспроизведения как часто встречающихся, так и редких слов.

E) Результаты одномерных тестов:

Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	V1	15.350(a)	3	5.117	2.883	.068
	V2	23.350(b)	3	7.783	9.434	.001
	V3	26.200(c)	3	8.733	11.644	.000

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	V1	281.250	1	281.250	158.451	.000
	V2	378.450	1	378.450	458.727	.000
	V3	369.800	1	369.800	493.067	.000
A	V1	4.050	1	4.050	2.282	.150
	V2	14.450	1	14.450	17.515	.001
	V3	16.200	1	16.200	21.600	.000
B	V1	11.250	1	11.250	6.338	.023
	V2	8.450	1	8.450	10.242	.006
	V3	9.800	1	9.800	13.067	.002
A * B	V1	.050	1	.050	.028	.869
	V2	.450	1	.450	.545	.471
	V3	.200	1	.200	.267	.613
Error	V1	28.400	16	1.775		
	V2	13.200	16	.825		
	V3	12.000	16	.750		
Total	V1	325.000	20			
	V2	415.000	20			
	V3	408.000	20			
Corrected Total	V1	43.750	19			
	V2	36.550	19			
	V3	38.200	19			

a R Squared = .351 (Adjusted R Squared = .229)
b R Squared = .639 (Adjusted R Squared = .571)
c R Squared = .686 (Adjusted R Squared = .627)

Результаты одномерных тестов детализируют статистически достоверные результаты многомерных тестов. Влияние фактора *А* проявляется статистически достоверно в отношении v2 — середины ряда и v3 — конца ряда. Судя по средним значениям для разных градаций этого фактора (см. **Descriptive Statistics** — Описательные статистики), при интонационном выделении середины ряда статистически достоверно увеличивается продуктивность воспроизведения середины ряда и уменьшается продуктивность воспроизведения конца ряда слов. Это различие наблюдается независимо от частоты встречаемости слов — как для часто встречающихся, так и для редких слов. Влияние фактора частоты встречаемости слов проявляется статистически достоверно в отношении всех трех зависимых переменных. Судя по средним значениям для разных градаций этого фактора, продуктивность воспроизведения редких слов ниже для каждой из трех частей ряда слов, чем продуктивность воспроизведения часто встречающихся слов.

Часть III

МНОГОМЕРНЫЕ МЕТОДЫ И МОДЕЛИ

Глава 14

НАЗНАЧЕНИЕ И КЛАССИФИКАЦИЯ МНОГОМЕРНЫХ МЕТОДОВ

При изучении основ применения математических методов акцент ставится на процедурах проверки статистических гипотез. В XX столетии статистическая проверка гипотез становится обязательной в науке вплоть до того, что логика статистического вывода начинает диктовать характер и форму самого исследования. При этом зачастую получается, что выявленные «статистически достоверные закономерности» с точки зрения содержательной теории или здравого смысла имеют в лучшем случае ничтожную ценность. Обилие подобных фактов вызывает справедливую критику абсолютизации теории статистического вывода в исследованиях, появляются даже крайние точки зрения, ставящие под сомнение саму возможность использования математических методов в психологии. Ведь действительно, выбор и использование статистических критериев не свободны от произвола исследователя: во многих случаях для одной и той же гипотезы можно подобрать статистический критерий, который ее отбросит, или тот критерий, который ее подтвердит. Остается уповать только на добросовестность исследователя.

При абсолютизации аппарата проверки гипотез забывают о том, что роль математических методов не только, да и не столько в статистическом обосновании предположений. Более значима *исходная функция математических методов в любой области знания — представление эмпирических данных в пригодном для интерпретации виде*, поиск смысла в обилии исходной информации. Ведь очень часто прежде, чем сформулировать гипотезу, мы «для себя» вычисляем средние значения, сравниваем частоты и т. д., то есть пытаемся осмыслить данные. Часто эти простейшие операции и не ассоциируются с применением математических методов. А на самом деле, с них и начинается использование математических методов в их основных назначениях — поисковых и описательных.

Здесь уместно ввести понятие *эмпирической математической модели* (ЭММ). Это *описательные математические модели*, применяемые для представления исходных (эмпирических) данных в доступном для интерпретации виде. Простейшие ЭММ — средние значения признака, вычисляемые для сравниваемых групп в предположении, что различия в средних значениях отражают

различия между представителями групп. Или даже просто ранжирование членов группы, которое предполагает, что порядковый номер испытуемого отражает выраженность изучаемого свойства. Если у нас два признака, измеренных на группе объектов (испытуемых), то мы вычисляем коэффициент корреляции или сопряженности, исходя из предположения о согласованности индивидуальной изменчивости признаков.

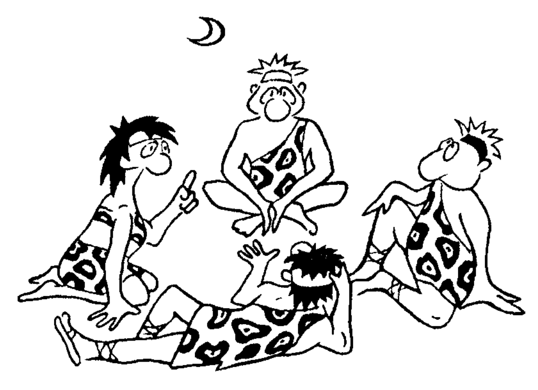
По сути дела, ЭММ идентичны мыслительным операциям. Но непосредственно сравнивать, различать, определять взаимосвязь и т. д. мы можем только при небольшой численности объектов или признаков. Когда объектов много, а признаков один-два, мы начинаем подсчитывать, если объектов более десятка — мы берем калькулятор. Когда много и объектов и признаков, простейшие ЭММ уже мало пригодны. И тогда возникает необходимость применения многомерных методов и компьютера.

Многомерные методы, таким образом, это дальнейшее развитие ЭММ в отношении многостороннего (многомерного) описания изучаемых явлений. Как и простейшие ЭММ, многомерные ЭММ воспроизводят мыслительные операции человека, но в отношении таких данных, непосредственное осмысление которых невозможно в силу нашей природной ограниченности. Многомерные методы выполняют такие интеллектуальные функции, как структурирование эмпирической информации (факторный анализ), классификация (кластерный анализ), экстраполяция (множественный регрессионный анализ), распознавание образов (дискриминантный анализ) и т. д.

Идея применения многомерных методов возникла практически одновременно с началом измерений в психологии. В конце XIX века Ф. Гальтон высказал мысль об общем факторе способностей, лежащем в основе согласованной индивидуальной изменчивости разных показателей способностей. Ч. Спирмен в начале XX века реализовал эту идею в однофакторном анализе для обоснования модели общего интеллекта. В 1930-е годы другой психолог, Л. Терстоун, предложил многофакторную математическую модель интеллекта и процедуру факторного анализа для ее верификации.

На протяжении последующих двух десятков лет факторный анализ не признавался математиками, ассоциируясь лишь с психологической моделью интеллекта. В 1950-е годы, с появлением ЭВМ, расширение применения факторного анализа в психологии сопровождается совершенствованием его математического аппарата. Итог этих лет — выход факторного анализа за пределы психологии, его внедрение в различные области знания в качестве популярного общенаучного метода. Другой итог — касается самой психологии: факторный анализ используется как основной инструмент при разработке наиболее известных тестовых методик (Р. Кеттелл, Г. Айзенк, Д. Векслер, Р. Амтхауэр и т. д.). Дальнейшее развитие психодиагностики складывается в соответствии с известным каноном: величие открытия определяется степенью последующего торможения развития науки в данном направлении. В настоящее время, несмотря на возникновение новых многомерных методов, в том числе в психологии (многомерное шкалирование), развитие психодиагностики буквально блокировано факторно-аналитичес-

кой традицией: мы измеряем признаки и сортируем их по факторам на основе взаимных корреляций.



С 1960-х годов, в связи с развитием компьютеризации, появляются все новые и новые методы многомерного анализа данных. Однако их широкое применение становится возможным лишь к концу 1980-х годов, с распространением персональных компьютеров. Дело в том, что любой многомерный метод требует циклической обработки данных, где на каждом этапе сам исследователь должен принимать решение о характере обработки. Поэтому раньше корректная реализация многомерного метода, например факторного анализа, требовала недель работы группы специалистов: предметника (психолога), статистика, программиста, оператора и др. Далеко не каждая исследовательская лаборатория могла себе это позволить.

В настоящее время, с появлением мощных и простых в применении программных средств, сам специалист может реализовать весь процесс многомерного анализа данных, не вдаваясь в вычислительные сложности. Для этого ему достаточно знать общий смысл метода, требования к исходным данным и основные показатели для интерпретации получаемых результатов. Вот этих-то знаний, компактных по объему и далеких от вычислительных тонкостей, ему часто и не хватает.

Наш опыт работы с многомерными методами начался с осознания необходимости создания удобного инструмента для их применения. Так, под нашим руководством в конце 1980-х годов была создана «Диалоговая система многомерного анализа экспериментальных данных» (ДИСМА — для ДВК, ISMEX — для PC). Работа этой программы на протяжении 10 лет в качестве основной для статистической обработки данных на факультете психологии СПбГУ выявила преобладание консервативной традиции у большинства специалистов. Понятие многомерных данных ограничивается факторным анализом, а сам анализ понимают как однократное преобразование корреляций в факторные нагрузки. Такое упрощенное понимание, помимо малоэффективного использования эмпирики, ограничивает перспективы развития как методов психодиагностики, так и методов организации исследования. Во-первых, любой многомерный анализ предполагает многократное преобразование данных, например факторный анализ — с разным числом факторов и с вариацией набора переменных. Во-вторых, факторный анализ — далеко не единственный, а часто и не оптимальный многомерный метод. В-третьих, ограниченность факторно-аналитической моделью задает слишком узкий диапазон способов организации самого эмпирического исследования: его данные должны представлять собой «множество признаков, измеренных у множества испытуемых».

Список многомерных методов, рассмотренных в этом разделе книги, не претендует на полноту: здесь изложены основы наиболее часто применяемых в психологии многомерных методов.

Представленные методы можно классифицировать по трем основаниям: в соответствии с интеллектуальной операцией (по способу преобразования исходной информации) — по назначению метода; по способу сопоставления данных — по сходству (различию) или пропорциональности (корреляции); по виду исходных эмпирических данных.

Классификация методов по назначению:

1. *Методы предсказания* (экстраполяции): множественный регрессионный и дискриминантный анализ. Множественный регрессионный анализ предсказывает значения метрической «зависимой» переменной по множеству известных значений «независимых» переменных, измеренных у множества объектов (испытуемых). Дискриминантный анализ предсказывает принадлежность объектов (испытуемых) к одному из известных классов (номинативной шкале) по измеренным метрическим (дискриминантным) переменным.

2. *Методы классификации*: варианты кластерного анализа и дискриминантный анализ. Кластерный анализ («классификация без обучения») по измеренным характеристикам у множества объектов (испытуемых) либо по данным об их попарном сходстве (различии) разбивает это множество объектов на группы, в каждой из которых содержатся объекты, более похожие друг на друга, чем на объекты из других групп. Дискриминантный анализ («классификация с обучением», «распознавание образов») позволяет классифицировать объекты по известным классам, исходя из измеренных у них признаков, пользуясь решающими правилами, выработанными предварительно на выборке идентичных объектов, у которых были измерены те же признаки.

3. *Структурные методы*: факторный анализ и многомерное шкалирование. Факторный анализ направлен на выявление структуры переменных как совокупности факторов, каждый из которых — это скрытая, обобщающая причина взаимосвязи группы переменных. Многомерное шкалирование выявляет шкалы как критерии, по которым поляризуются объекты при их субъективном попарном сравнении.

Классификация методов по исходным предположениям о структуре данных:

1. Методы, исходящие из предположения о согласованной изменчивости признаков, измеренных у множества объектов: факторный анализ, множественный регрессионный анализ, отчасти — дискриминантный анализ.

2. Методы, исходящие из предположения о том, что различия между объектами можно описать как расстояние между ними. На *дистантной модели* основаны кластерный анализ и многомерное шкалирование, частично — дискриминантный анализ. Многомерное шкалирование и дискриминантный анализ добавляют предположение о том, что исходные различия между объектами можно представить как расстояния между ними в пространстве небольшого числа шкал (функций).

Классификация методов по виду исходных данных:

1. Методы, использующие в качестве исходных данных только признаки, измеренные у группы объектов. Это множественный регрессионный анализ, дискриминантный анализ и факторный анализ.

2. Методы, исходными данными для которых могут быть попарные сходства (различия) между объектами: это кластерный анализ и многомерное шкалирование. Многомерное шкалирование, кроме того, может анализировать данные о попарном сходстве между совокупностью объектов, оцененном группой экспертов. При этом совместно анализируются как различия между объектами, так и индивидуальные различия между экспертами.

Представленные классификации свидетельствуют о необходимости знаний многомерных методов, их возможностей и ограничений уже на стадии общего замысла исследования. Например, ориентируясь только на факторно-аналитическую модель, исследователь ограничен в выборе процедуры диагностики: она должна состоять в измерении признаков у множества объектов. При этом исследователь ограничен и в направлении поиска: он изучает либо взаимосвязи между признаками, либо межгрупповые различия по измеряемым признакам. Общая осведомленность о других многомерных методах позволит исследователю использовать более широкий круг психодиагностических процедур, решать более широкий спектр не только научных, но и практических задач.

Применение многомерных методов требует, разумеется, не только самого компьютера, но и соответствующего программного обеспечения. Широко известны и распространены универсальные статистические программы STATISTICA и SPSS, содержащие практически весь спектр статистических методов — от простейших до самых современных. Мы разделяем мнение, что программа STATISTICA обладает прекрасной графикой и гибкостью в обработке данных. Однако программа SPSS имеет свои преимущества: она не только проще в освоении и применении, но и включает в себя ряд методов, отсутствующих в STATISTICA, например, варианты многомерного шкалирования.

Глава 15

МНОЖЕСТВЕННЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

НАЗНАЧЕНИЕ

Множественный регрессионный анализ (МРА) предназначен для изучения взаимосвязи одной переменной (*зависимой, результирующей*) и нескольких других переменных (*независимых, исходных*)[1]. Исходные данные для МРА представляют собой таблицу (матрицу) размерностью $N \times P$ следующего вида:

№	X_1	X_2	...	X_P
1	x_{11}	x_{12}	...	x_{1P}
2	x_{21}	x_{22}	...	x_{2P}
...	...	...	...	...
N	x_{N1}	x_{N2}	...	x_{NP}

Строки этой таблицы соответствуют объектам (испытуемым), а столбцы — переменным. Все переменные при этом должны быть измерены в количественной шкале. Одна из переменных определяется исследователем как зависимая, а остальные (или часть их) — как независимые переменные. Допускается, что для некоторых объектов значения зависимой переменной неизвестны, и их определение (оценка) может составлять важный результат анализа.

МРА может применяться как для решения прикладных задач, так и в исследовательских целях. Обычно МРА применяется для изучения возможности предсказания некоторого результата (обучения, деятельности) по ряду предварительно измеренных характеристик. При этом предполагается, что связь между одной зависимой переменной (Y) и несколькими независимыми переменными (X) можно выразить линейным уравнением:

$$Y = b + b_1x_1 + b_2x_2 + ... + b_Px_P + e, \quad (15.1)$$

где Y — **зависимая переменная**; $x_1, ..., x_P$ — **независимые переменные**; $b, b_1, ..., b_P$ — параметры модели; e — ошибка предсказания.

[1] Перед чтением этой главы рекомендуется освежить в памяти материал главы 6 о корреляции и двумерной регрессии.

ПРИМЕРЫ

Психолога может заинтересовать предсказание успеваемости абитуриента по измеренным психологическим характеристикам (интеллекта, личности и пр.). В этом случае он использует уже имеющиеся данные о взаимосвязи успеваемости и предварительного психологического тестирования за прошлые годы. Успеваемость при этом он рассматривает как зависимую переменную, психологические показатели — как независимые переменные. Применяя МРА, он получает модель предсказания в виде 15.1. Подставляя в эту модель данные абитуриента, психолог получает предсказание его успеваемости.

Сходным образом психолог может изучать удовлетворенность оплатой труда. Привлекая данные разных компаний, он может при помощи МРА определить зависимость оплаты труда (Y) сотрудника от степени ответственности, количества подчиненных и других показателей ($x_1, ..., x_P$). Пользуясь этой моделью, можно определить сотрудников, которым недоплачивают, переплачивают или платят «справедливо» за их труд.

Р. Кеттелл при помощи МРА получил «профессиональные портреты» для некоторых специальностей:

- психотерапевт = $0{,}72A + 0{,}29B + 0{,}29H + 0{,}29N$;
- психодиагност = $0{,}31A + 0{,}78B + 0{,}47N$.

Коэффициенты регрессии перед сокращенными техническими обозначениями шкал-факторов опросника Р. Кеттелла указывают на их вклад в прогноз эффективности соответствующей деятельности. Так, для психотерапевта важнее всего общительность (A), а для психодиагноста — интеллект (B).

Помимо предсказания и определения степени его точности МРА позволяет определить и то, какие показатели («независимые переменные») наиболее существенны, важны для предсказания, а какими переменными можно пренебречь, исключив их из анализа. Например, психолога может интересовать вопрос о том, какие психологические характеристики в наибольшей степени влияют на проявление исследуемой формы поведения или какие индивидуальные особенности лучше предсказывают успешность деятельности и пр.

Следует отметить родственность множественного регрессионного и дисперсионного анализа. В основе этих методов лежит одна и та же *линейная модель* 15.1. Этот факт отражает и название, которое объединяет различные варианты дисперсионного анализа (в частности, в программе SPSS): **общая линейная модель** (*General Linear Model*). МРА в этом смысле можно рассматривать как *аналог многофакторного дисперсионного анализа* для случая, когда независимые переменные представляют собой не градации факторов (номинативные переменные), а измерены в количественной шкале. Тогда, в соответствии с моделью 15.1, МРА выступает как инструмент исследования влияния факторов (независимых переменных) $x_1, ..., x_P$ на зависимую переменную Y.

Часто зависимая переменная Y выступает в качестве градаций, которым соответствуют разные группы объектов, т. е. измерена в номинативной шкале. В этом случае модель множественной регрессии неприемлема, и вместо МРА может быть применен *дискриминантный анализ*, который решает те же задачи и позволяет получить сходные результаты (см. главу 16).

МРА может применяться и в том случае, если переменная Y является причиной изменения нескольких переменных $x_1, ..., x_p$. Так, зависимой переменной может быть скрытая причина, фактор, например личностное свойство, а независимыми переменными — пункты теста, измеряющие различные проявления этого свойства. Таким образом, понятия «зависимая» и «независимая» переменные в МРА являются условными, а определение направления причинно-следственной связи выходит за рамки применения самого метода.

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ ИДЕИ МЕТОДА

Исходным положением линейного МРА является возможность представления значений «зависимой» переменной Y через значения «независимых» переменных $x_1, x_2, ..., x_p$ в виде линейного уравнения:

$$Y = b + b_1x_1 + b_2x_2 + ... + b_px_p + e,$$

где b — **свободный член** (*Intercept*), $b_1, ..., b_p$ — b — **коэффициенты регрессии** (*Unstandardized Coefficients*), e — **ошибка оценки** (*Residual*). Коэффициенты регрессии вычисляются методом наименьших квадратов при решении системы из линейных уравнений, с минимизацией ошибки e.

После вычисления регрессионных коэффициентов по значениям независимых переменных для каждого из объектов могут быть вычислены **оценки зависимой переменной** Y (*Predicted Values*):

$$\hat{Y} = b + b_1x_1 + b_2x_2 + ... + b_px_p. \quad (15.2)$$

Сопоставление значений зависимой переменной Y_i с их оценками $\hat{Y}_i$ по выборке испытуемых, для которых значения Y_i известны, называется анализом остатков или ошибок (*residual analysis*). Он позволяет оценить возможные погрешности предсказания. Значения оценок $\hat{Y}_i$ могут быть вычислены и для испытуемых, истинные значения зависимой переменной для которых неизвестны.

Далее можно вычислить коэффициент корреляции Пирсона между известными значениями «зависимой» переменной и ее оценками. Это один из способов получения *коэффициента множественной корреляции* (КМК) между «зависимой» и «независимыми» переменными. **Коэффициент множественной корреляции** — это мера линейной связи одной переменной с множеством других переменных; принимает положительные значения от 0 (отсутствие связи) до 1 (строгая прямая связь). КМК наряду с разностями между исходными и оцененными значениями «зависимой» переменной (ошибки e) — основные показатели качества модели множественной регрессии.

Если «зависимая» и «независимые» переменные представлены в z-значениях, то уравнение регрессии принимает вид:

$$Y_z = \beta_1x_1 + \beta_2x_2 + ... + \beta_px_p + e, \quad (15.3)$$

где β_P — **стандартные коэффициенты регрессии**, или **β-коэффициенты** (*Standardized Coefficients*).

Стандартные коэффициенты регрессии связаны с исходными корреляциями следующим уравнением (в матричной форме):

$$B = R^{-1}A, \qquad (15.4)$$

где B — вектор-столбец стандартных коэффициентов регрессии, R^{-1} — матрица, обратная корреляционной матрице «независимых» переменных, A — вектор-столбец корреляций «независимых» переменных с «зависимой» переменной. На практике регрессионный анализ начинается именно с вычисления стандартных коэффициентов регрессии.

Напомним, что в случае двумерной регрессии — при наличии всего одной независимой переменной, уравнение 15.3 имеет вид:

$$\hat{y}_i = r_{xy} \cdot x_i\,,$$

то есть стандартный коэффициент регрессии равен коэффициенту корреляции зависимой и независимой переменных. При наличии двух и более независимых переменных:

$$|\beta_x| \leqslant |r_{xy}|$$

и β-коэффициент зависит не только от корреляции данной независимой и зависимой переменных, но и от того, коррелирует ли эта независимая переменная с другими независимыми переменными. Знак β-коэффициента соответствует знаку коэффициента корреляции данной «независимой» и «зависимой» переменной. Абсолютная величина β-коэффициента является максимальной — равна коэффициенту корреляции с зависимой переменной, если данная независимая переменная не коррелирует ни с одной из других независимых переменных. Чем сильнее данная независимая переменная связана с другими независимыми переменными, тем меньше β-коэффициент.

Произведение коэффициента β_i на коэффициент корреляции r_{iY} этой переменной с «зависимой» переменной — это *вклад* данной переменной в дисперсию «зависимой» переменной. Ясно, что вклад переменной выше, если ее корреляция с зависимой переменной выше, а с другими независимыми переменными — ниже. Поэтому ценность независимой переменной для множественной регрессии определяется не только ее корреляцией с зависимой переменной (как в двумерной регрессии), но и ее «уникальностью» — слабой связью с другими независимыми переменными.

Если «зависимая» переменная представлена в z-значениях (дисперсия равна 1), то эта единичная дисперсия «зависимой» переменной D_Y может быть выражена формулой:

$$D_Y = 1 = \sum_{i=1}^{P} \beta_i r_{iY} + D_e\,,$$

где D_e — часть дисперсии, обусловленная влиянием неучтенных факторов, или *дисперсия ошибки предсказания*.

Часть дисперсии «зависимой» переменной, обусловленная влиянием «независимых» переменных, — это **коэффициент множественной детерминации** (КМД), который равен коэффициенту множественной корреляции в квадрате или R^2:

$$\text{КМД} = R^2 = \sum_{i=1}^{P} \beta_i r_{iY} = 1 - D_e.$$

Соответственно, второй способ вычленить КМК:

$$R = \sqrt{\sum_{i=1}^{P} \beta_i r_{iY}}\,.$$

Интерпретация КМД очевидна: это та часть дисперсии «зависимой» переменной, которая определяется «независимыми» переменными. Следовательно, (1 – КМД) — это дисперсия ошибки оценки. Например, если КМК = 0,8, то КМД = (КМК)^2 = 0,64. Это означает, что 64% дисперсии «зависимой» переменной определяется исходными переменными, а 36% ее дисперсии относится к ошибке оценки.

Таким образом, основной показатель МРА — *коэффициент множественной корреляции* (R), который, подобно парному коэффициенту корреляции Пирсона, является мерой линейной взаимосвязи одной переменной с совокупностью других переменных. КМК «зависимой» переменной с набором «независимых» переменных, как и КМД, принимает только положительные значения, изменяясь в пределах от 0 до 1. Статистическая значимость КМК определяется по критерию F-Фишера для соответствующих степеней свободы.

Таким образом, основными целями МРА являются:

1. Определение того, в какой мере «зависимая» переменная связана с совокупностью «независимых» переменных, какова статистическая значимость этой взаимосвязи. Показатель — коэффициент множественной корреляции (КМК) и его статистическая значимость по критерию F-Фишера.

2. Определение существенности вклада каждой «независимой» переменной в оценку «зависимой» переменной, отсев несущественных для предсказания «независимых» переменных. Показатели — регрессионные коэффициенты β, их статистическая значимость по критерию t-Стьюдента.

3. Анализ точности предсказания и вероятных ошибок оценки «зависимой» переменной. Показатель — квадрат КМК, интерпретируемый как доля дисперсии «зависимой» переменной, объясняемая совокупностью «независимых» переменных. Вероятные ошибки предсказания анализируются по расхождению (разности) действительных значений «зависимой» переменной и оцененных при помощи модели МРА.

4. Оценка (предсказание) неизвестных значений «зависимой» переменной по известным значениям «независимых» переменных. Осуществляется по вычисленным параметрам множественной регрессии.

ИСХОДНЫЕ ДАННЫЕ, ПРОЦЕДУРА И РЕЗУЛЬТАТЫ

Исходными данными для МРА является набор переменных, измеренных для выборки объектов (испытуемых). Одна из переменных определяется как «зависимая», остальные — как «независимые» переменные.

ПРИМЕР 15.1

Перед исследователем стоит задача предсказания успеваемости пяти абитуриентов по данным вступительных тестов (4 теста). Кроме того, его интересует, какие тесты обладают наибольшей предсказательной силой в отношении последующей успеваемости. В качестве исходных данных психолог имеет для каждого из 20 учащихся предыдущего набора средний балл отметок и 4 показателя тестирования. В его распоряжении имеются результаты применения тех же 4 тестов для пяти абитуриентов, и исследователь надеется предсказать для них средний балл успеваемости. Таким образом, исходными данными для МРА являются: средний балл отметок как «зависимая» переменная (Y) и 4 «независимых» переменных — результатов тестов (test 1, test 2, test 3, test 4) (табл. 15.1).

Таблица 15.1

Пример исходных данных для МРА

№	test 1	test 2	test 3	test 4	Y	$\hat{Y}$
1	86,00	110,00	110,00	101,00	3,88	
2	80,00	97,00	99,00	100,00	3,64	
3	93,00	107,00	103,00	103,00	4,11	
4	87,00	117,00	93,00	88,00	3,54	
...	...	...	...	...	...	
20	120,00	94,00	110,00	105,00	3,71	
21	74,00	121,00	100,00	100,00		
22	96,00	114,00	114,00	103,00		
23	104,00	73,00	105,00	95,00		
24	94,00	121,00	115,00	104,00		
25	91,00	129,00	105,00	98,00		

Первые 20 объектов — это учащиеся предыдущего набора, для которых известен средний балл успеваемости, последние 5 объектов — это абитуриенты, для которых известны только результаты тестирования. Последний столбец ($\hat{Y}$) — это оценки «зависимой» переменной, которые исследователь надеется получить в результате применения МРА. Корреляции исходных переменных приведены в табл. 15.2.

Таблица 15.2

Корреляция исходных данных для МРА

	test 1	test 2	test 3	test 4	Y
test 1	1	–0,015	0,263	0,402	0,639
test 2	–0,015	1	0,356	0,317	0,552
test 3	0,263	0,356	1	0,772	0,706
test 4	0,402	0,317	0,772	1	0,736
Y	0,639	0,552	0,706	0,736	1

Строгих указаний о соотношении количества объектов N и количества признаков P нет, но чем больше объем выборки, тем выше шансы получить статистически достоверные результаты.

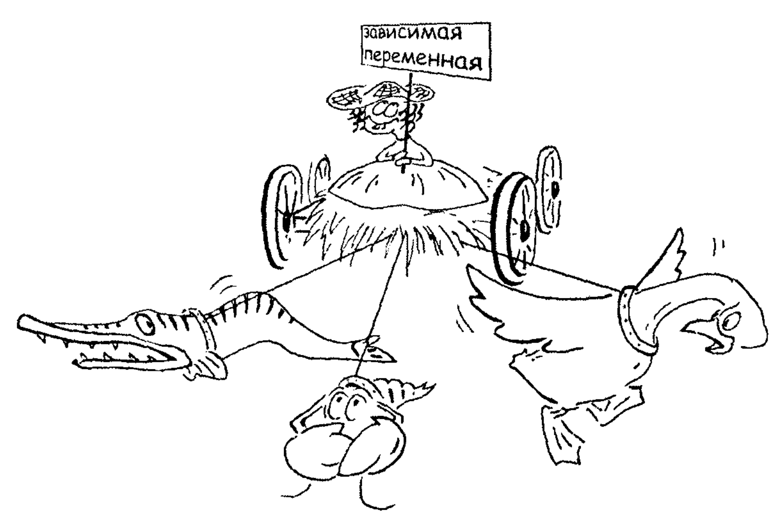
Главное требование к исходным данным — отсутствие линейных взаимосвязей между переменными, когда одна переменная является линейной производной другой переменной. Таким образом, нельзя пользоваться суммой переменных или их средним арифметическим наряду с самими переменными. Соответственно, недопустимы переменные, коэффициент корреляции которых с любой другой переменной равен 1. Следует избегать включения в анализ переменных, корреляция между которыми близка к 1, так как сильно коррелирующая переменная не несет для анализа новой информации, добавляя излишний «шум».

Следующее требование — переменные должны быть измерены в метрической шкале (интервалов или отношений) и иметь нормальное распределение. При нарушении этого требования, однако, результаты могут быть полезны, если, конечно, соблюдать известную осторожность.

Желательно отбирать для МРА «независимые» переменные, сильно коррелирующие с «зависимой» переменной и слабо — друг с другом. Если «независимых» переменных много и наблюдается множество связей между ними, то перед МРА целесообразно провести факторный анализ этих «независимых» переменных с вычислением значений факторов для объектов (см. главу 16).

Практически во всех компьютерных программах анализ начинается с задания «зависимой» и «независимых» переменных, а также одного из методов МРА. *Основные методы МРА:*

- исходный или стандартный (Enter);
- прямой пошаговый (Forward);
- обратный пошаговый (Backward).



Независимые переменные не должны быть связаны друг с другом

Стандартный метод учитывает в МРА все «зависимые» переменные. Пошаговый метод обычно выступает в нескольких модификациях, основными из которых являются прямой и обратный метод.

Прямой пошаговый метод поочередно включает в регрессионное уравнение каждую переменную, начиная с наиболее тесно коррелирующей с «зависимой» переменной, до тех пор, пока p-уровень значимости β-коэффициента последней из включенных переменных не превысит заданное значение (по умолчанию — 0,1). **Обратный пошаговый метод** поочередно исключает переменные из анализа, начиная с той, которая имеет наибольшее значение p-уровня значимости β-коэффициента, до тех пор, пока все оставшиеся переменные не будут иметь статистически значимые β-коэффициенты (по умолчанию $p \leqslant 0{,}1$). Таким образом, пошаговые методы позволяют отсеивать несущественные для предсказания «независимые» переменные — те, β-коэффициенты которых статистически не достоверны. Следует отметить, что разные варианты пошагового метода могут давать разные результаты, поэтому следует применить каждый из них и выбрать наиболее приемлемый конечный результат.

Основные результаты применения МРА:

R — коэффициент множественной корреляции;

F — критерий Фишера и p-уровень статистической значимости КМК;

R^2 — квадрат КМК или КМД;

β (*Beta*) — стандартизированные коэффициенты регрессии и p-уровень их статистической значимости;

B — коэффициенты регрессии (регрессионного уравнения).

Дополнительно возможно вычисление оценок «зависимой» переменной (Predicted Values) и ошибок оценки (Residuals).

Рассмотрим процедуру обработки, основные результаты и их интерпретацию, применив программу SPSS к данным примера 15.1.

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Исходные данные (**Data Editor**) представляют собой 5 переменных (`test 1`, `test 2`, `test 3`, `test 4`, `Y`) для 25 объектов; для последних 5 объектов значения Y не определены (табл. 15.1).

1. Выбираем **Analyze > Regression** (Регрессионный) > **Linear...** (Линейный).

2. В открывшемся основном окне диалога **Linear Regression** (Линейный регрессионный) выделяем и переносим из левого окна переменные при помощи кнопки ▷: зависимую переменную (Y) в правое верхнее окно (**Dependent**), независимые переменные (`test 1`, `test 2`, `test 3`, `test 4`) — в правое второе сверху окно (**Independents**).

3. В том же окне диалога выбираем метод. Для этого в окне **Method** (Метод) вместо принятого по умолчанию стандартного метода (**Enter**) при помо-

щи кнопки ▷ выбираем один из пошаговых методов, в данном случае — **Backward** (Обратный).

4. Для вычисления и сохранения оценок зависимой переменной[1] в том же окне диалога нажимаем клавишу **Save...** (Сохранить). В появившемся окне диалога в разделе **Predicted Values** (Предсказанные оценки) отмечаем флажком **Unstandardized** (Не стандартизованные). Нажимаем **Continue** (Продолжить).

5. После указания всех установок в основном окне диалога **Linear Regression** (Линейный регрессионный) нажимаем OK и получаем результаты.

Рассмотрим наиболее важные результаты МРА.

Model Summary(c)

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.886(a)	.786	.729	.29023
2	.879(b)	.773	.731	.28914

a Predictors: (Constant), TEST4, TEST2, TEST1, TEST3
b Predictors: (Constant), TEST4, TEST2, TEST1
c Dependent Variable: Y

ANOVA(c)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4.635	4	1.159	13.755	.000(a)
	Residual	1.264	15	.084		
	Total	5.898	19			
2	Regression	4.560	3	1.520	18.183	.000(b)
	Residual	1.338	16	.084		
	Total	5.898	19			

a Predictors: (Constant), TEST4, TEST2, TEST1, TEST3
b Predictors: (Constant), TEST4, TEST2, TEST1
c Dependent Variable: Y

Эти две таблицы содержат наиболее общие результаты МРА для двух моделей (Model): 1 — исходная модель, с включением всех переменных; 2 — окончательная модель, с исключенной переменной `test3`. Интерпретации подлежат: в первой таблице — КМК (R) и КМД (R Square); во второй таблице — значение критерия *F*-Фишера и его *p*-уровень значимости. КМК для окончательной модели статистически достоверен, поэтому модель множественной регрессии может быть содержательно интерпретирована. КМД достаточно большой, регрессионная модель объясняет более 77% дисперсии зависимой переменной, и результаты предсказания могут быть приняты во внимание.

[1] Этот шаг следует выполнять только на заключительном этапе проведения МРА, когда после предварительных проб определена окончательная модель регрессии.

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-1.486	.871		-1.705	.109
	TEST1	.011	.004	.384	2.825	.013
	TEST2	.012	.005	.334	2.565	.022
	TEST3	.014	.015	.196	.938	.363
	TEST4	.017	.012	.295	1.380	.188
2	(Constant)	-1.049	.733		-1.430	.172
	TEST1	.011	.004	.379	2.800	.013
	TEST2	.014	.005	.368	2.949	.009
	TEST4	.026	.008	.445	3.161	.006

a Dependent Variable: Y

Эта таблица содержит величины не стандартизованных (*B*) и стандартизованных (Beta) коэффициентов регрессии, а также критерии *t*-Стьюдента (*t*) и *p*-уровни (Sig.), позволяющие определить их статистическую значимость. Отметим, что во внимание могут быть приняты только те регрессионные коэффициенты, которые являются статистически значимыми.

Показателем вклада каждой из переменных в регрессионную модель служат их β-коэффициенты. В результирующей модели (2) для предсказания остаются три переменные: переменная test 3 исключена. Представляет интерес анализ причин исключения переменной test3 из модели. Как видно из табл. 15.2, эта переменная весьма существенно коррелирует с зависимой переменной. Однако она сильно связана и с переменной test 4, что обусловливает существенное снижение ее β-коэффициента: в исходной модели (1) он наименьший из всех остальных ($\beta = 0{,}196$). Исключение переменной test 3 повышает предсказательную ценность переменной test 4.

B-коэффициенты используются для предсказания значений зависимой переменной — путем вычисления ее оценок по уравнению регрессии, в соответствии с формулой 15.2:

$$\hat{Y} = -1{,}049 + 0{,}026(\text{test } 4) + 0{,}011(\text{test } 1) + 0{,}014(\text{test } 2).$$

Фрагмент таблицы (Data Editor) с оценками зависимой переменной ($\hat{Y}$)

	test 1	test 2	test 3	test 4	*Y*	pre_1
1	86,00	110,00	110,00	101,00	3,88	3,98
...	...	...	...	...	...	...
21	74,00	121,00	100,00	100,00		3,98
22	96,00	114,00	114,00	103,00		4,19
23	104,00	73,00	105,00	95,00		3,52
24	94,00	121,00	115,00	104,00		4,29
25	91,00	129,00	105,00	98,00		4,22

В связи с тем, что назначено сохранение оценок зависимой переменной (см. шаг 4), в таблице исходных данных (**Data Editor**) появилась новая переменная `pre_1` — оценки зависимой переменной. Они вычислены по указанной формуле для всех 25 объектов и в данном примере могут служить для предсказания успеваемости 5 абитуриентов.

Глава 16

ФАКТОРНЫЙ АНАЛИЗ

НАЗНАЧЕНИЕ

Возникновение и развитие факторного анализа тесно связано с измерениями в психологии. Длительное время факторный анализ и воспринимался как математическая модель в психологической теории интеллекта. Лишь начиная с 50-х годов XX столетия, одновременно с разработкой математического обоснования факторного анализа, этот метод становится общенаучным. К настоящему времени факторный анализ является неотъемлемой частью любой серьезной статистической компьютерной программы и входит в основной инструментарий всех наук, имеющих дело с многопараметрическим описанием изучаемых объектов, таких, как социология, экономика, биология, медицина и другие.

Основная идея факторного анализа была сформулирована еще Ф. Гальтоном, основоположником измерений индивидуальных различий. Она сводится к тому, что если несколько признаков, измеренных на группе индивидов, изменяются согласованно, то можно предположить существование одной общей причины этой совместной изменчивости — фактора как скрытой (латентной), непосредственно не доступной измерению переменной. Далее К. Пирсон в 1901 году выдвигает идею «метода главных осей», а Ч. Спирмен, отстаивая свою однофакторную концепцию интеллекта, разрабатывает математический аппарат для оценки этого фактора, исходя из множества измерений способностей. В своей работе, опубликованной в 1904 году, Ч. Спирмен показал, что если ряд признаков попарно коррелируют друг с другом, то может быть составлена система линейных уравнений, связывающих все эти признаки, один общий фактор «общей одаренности» и по одному специфическому фактору «специальных способностей» для каждой переменной. В 1930-х годах Л. Тер-



Фактор — скрытая причина согласованной изменчивости наблюдаемых переменных

стоун впервые предлагает «многофакторный анализ» для описания многочисленных измеренных способностей меньшим числом общих факторов интеллекта, являющихся линейной комбинацией этих исходных способностей.

С 1950-х годов, с появлением компьютеров, факторный анализ начинает очень широко использоваться в психологии при разработке тестов, обоснования структурных теорий интеллекта и личности. При этом исследователь начинает с множества измеренных эмпирических показателей, которые при помощи факторного анализа группируются по факторам (изучаемым свойствам). Факторы получают интерпретацию по входящим в них переменным, затем отбираются наиболее «весомые» показатели этих факторов, отсеиваются малозначимые переменные, вычисляются значения факторов для испытуемых и сопоставляются с внешними эмпирическими показателями изучаемых свойств.

В дальнейшем, по мере развития математического обеспечения факторного анализа, накопления опыта его использования, прежде всего в психологии, задача факторного анализа обобщается. Как общенаучный метод, факторный анализ становится средством для замены набора коррелирующих измерений существенно меньшим числом новых переменных (факторов). При этом основными требованиями являются: а) минимальная потеря информации, содержащейся в исходных данных, и б) возможность представления (интерпретации) факторов через исходные переменные.

Таким образом, *главная цель факторного анализа* — уменьшение размерности исходных данных с целью их экономного описания при условии минимальных потерь исходной информации. *Результатом* факторного анализа является переход от множества исходных переменных к существенно меньшему числу новых переменных — факторов. **Фактор** при этом интерпретируется как причина совместной изменчивости нескольких исходных переменных.

Если исходить из предположения о том, что корреляции могут быть объяснены влиянием скрытых причин — факторов, то *основное назначение* факторного анализа — *анализ корреляций* множества признаков.

ПРИМЕР 16.1

Рассмотрим результаты факторного анализа на простом примере. Предположим, исследователь измерил на выборке из 50 испытуемых 5 показателей интеллекта: счет в уме, продолжение числовых рядов, осведомленность, словарный запас, установление сходства. Все показатели статистически значимо взаимосвязаны на уровне $p < 0{,}05$, кроме показателя № 4 с № 1 и 2 (табл. 16.1).

Таблица 16.1

Матрица корреляций пяти показателей интеллекта

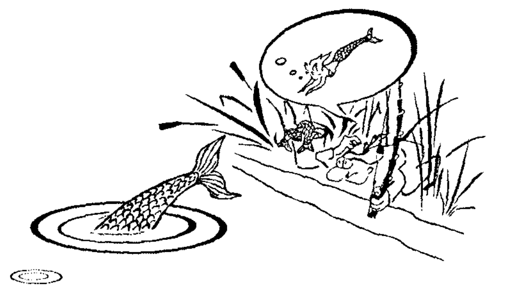
№	Показатели	1	2	3	4	5
1	Счет в уме	1,00	0,88	0,33	0,23	0,42
2	Числовые ряды	0,88	1,00	0,32	0,24	0,35
3	Осведомленность	0,33	0,32	1,00	0,58	0,58
4	Словарный запас	0,23	0,24	0,58	1,00	0,54
5	Сходство	0,42	0,35	0,58	0,54	1,00

Таблица 16.2

Факторные нагрузки после варимакс-вращения

Исходные переменные	Факторные нагрузки		h_2 (общность)
	F_1	F_2	
1	0,97	0,20	0,99
2	0,86	0,20	0,78
3	0,18	0,76	0,62
4	0,09	0,74	0,56
5	0,26	0,69	0,55
Собственное значение	1,79	1,70	3,5
Доля дисперсии	0,36	0,34	0,7

Применив факторный анализ, исследователь выделил два фактора. Основной результат, который подлежит интерпретации исследователем, — таблица факторных нагрузок после варимакс-вращения (табл. 16.2). Не рассматривая пока шаги, приводящие к этому результату, попытаемся проинтерпретировать полученные данные. В нашем примере по фактору 1 (F_1) максимальные нагрузки имеют переменные 1 и 2. Следовательно, фактор 1 и определяется этими переменными. Поскольку переменная 1 — счет в уме, а переменная 2 — продолжение числового ряда, то фактору 1 может быть присвоено название «арифметические способности», как показателю легкости оперирования числовым материалом. Точно так же фактору 2 можно присвоить название «вербальные способности», как показателю словесного понимания. Нетрудно заметить, что переменные, определяющие фактор, сильнее связаны друг с другом, чем с другими переменными (табл. 16.1). Так, переменные 1 и 2, определяющие фактор 1, сильнее связаны друг с другом, чем с переменными 3, 4 и 5. Таким образом, за взаимосвязью пяти исходных измерений способностей при помощи факторного анализа обнаруживается действие двух латентных переменных (факторов).



Интерпретация фактора через исходные переменные

Интерпретация факторов — одна из основных задач факторного анализа. Ее решение заключается в идентификации факторов через исходные переменные. Эта идентификация и осуществляется по результатам обработки, представленным в табл. 16.2.

Основное содержание табл. 16.2 — величины $a_{11} \ldots a_{25}$ — *факторные нагрузки* переменных 1 ... 5 (строки) по факторам 1 и 2 (столбцы). Факторные нагрузки — аналоги коэффициентов корреляции, показывают степень взаимосвязи соответствующих переменных и факторов: чем больше *абсолютная* величина факторной нагрузки, тем сильнее связь переменной с фактором, тем больше данная переменная обусловлена действием соответствующего фактора. Каждый фактор идентифицируется по тем переменным, с которыми он в наибольшей степени связан, то есть по переменным, имеющим по

этому фактору наибольшие нагрузки. Идентификация фактора заключается, как правило, в присвоении ему имени, обобщающего по смыслу наименования входящих в него переменных.

Если исследователя интересует только структура измеренных признаков, на этом факторный анализ завершается. Продолжая факторный анализ, исследователь далее может вычислить значения факторов для испытуемых, например, с целью их дифференциации по преобладанию арифметических или вербальных способностей.

Выбирая факторный анализ как средство изучения корреляций, исследователь должен отдавать себе отчет в том, что это один из самых сложных и трудоемких методов. Зачастую нет веских оснований предполагать наличие факторов как скрытых причин изучаемых корреляции, и задача заключается лишь в обнаружении группировок тесно связанных переменных. Тогда целесообразнее вместо факторного анализа использовать **кластерный анализ корреляций** (см. главу 19). Помимо простоты, кластерный анализ обладает еще одним преимуществом: его применение не связано с потерей исходной информации о связях между переменными, что неизбежно при факторном анализе. И уже после выделения групп тесно связанных переменных можно попытаться применить факторный анализ для их объяснения.

Итак, можно сформулировать основные задачи факторного анализа:

1. Исследование структуры взаимосвязей переменных. В этом случае каждая группировка переменных будет определяться фактором, по которому эти переменные имеют максимальные нагрузки.

2. Идентификация факторов как скрытых (латентных) переменных — причин взаимосвязи исходных переменных.

3. Вычисление значений факторов для испытуемых как новых, интегральных переменных. При этом число факторов существенно меньше числа исходных переменных. В этом смысле факторный анализ решает задачу сокращения количества признаков с минимальными потерями исходной информации.

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ ИДЕИ И ПРОБЛЕМЫ МЕТОДА

Анализ главных компонент и факторный анализ

Модель главных компонент лежит в основе большинства методов факторного анализа и часто рассматривается как один из его самостоятельных вариантов. **Анализ главных компонент** преобразует набор коррелирующих исходных переменных в другой набор — некоррелирующих переменных. Проще всего понять суть этого метода, привлекая геометрические представления.

Предположим, у нас имеются две положительно коррелирующие переменные X и Y, измеренные на группе объектов. Тогда график двумерного распределения (рассеивания) этих объектов в осях измеренных признаков (координаты объектов заданы значениями признаков) будет представлять собой эллипс, так как большим значениям переменной X будут соответствовать большие значения переменной Y и наоборот (рис. 16.1). Главная ось эллипса M_1 — это прямая, вдоль которой будет наблюдаться наибольший разброс данных. Вдоль второй оси эллипса M_2, перпендикулярной первой и проходящей через ее середину, будет наблюдаться наименьший разброс данных.

Если перед нами стоит задача представления объектов (точек) в терминах только одной размерности (переменной), то главная ось эллипса является наиболее подходящей, так как вдоль нее объекты отличаются друг от друга лучше (дисперсия больше), чем вдоль любой другой прямой, в том числе и вдоль отдельно оси X или Y. Анализ главных компонент в отношении этих двух признаков и состоит в переходе от них к главной компоненте, соответствующей главной оси эллипса, и в представлении объектов в значениях проекций объектов на эту ось (*главную компоненту*). Иначе говоря, происходит переход от координат каждого объекта по двум осям (X, Y) к их координатам только по одной оси M_1 — главной компоненте (рис. 16.1). Отметим, что в случае отсутствия взаимосвязи двух признаков главной компоненты просто не существует, так как обе оси (компоненты) являются равнозначными.

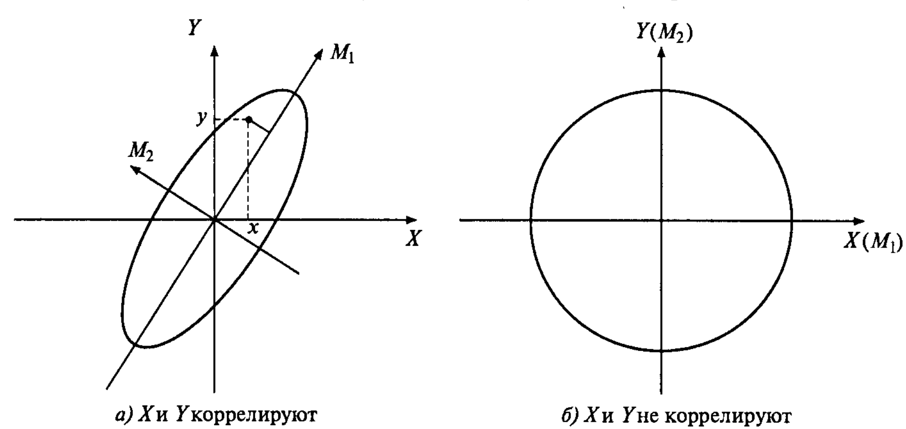


Рис. 16.1. Компоненты M_1 и M_2 двумерного распределения признаков X и Y

Анализ главных компонент можно представить как преобразование информации, содержащейся в исходных данных. Так, определяя главную компоненту как направление, в котором наблюдается наибольший разброс объектов, представляя объекты в единицах измерения по этой оси, мы теряем минимум информации об отличии объектов друг от друга. Чем сильнее взаимосвязь двух переменных, тем меньше исходной информации теряется при переходе от двух переменных к одной главной компоненте. Если две переменные не коррелируют, то компоненты (оси) являются равнозначными по информативности, и невозможно определить одну из них как «главную».

При наличии более двух коррелирующих переменных принцип определения главных компонент тот же. В осях трех и более переменных график разброса объектов будет представлять собой эллипсоид (овальное тело) в пространстве трех и более измерений. Первая ось этого эллипсоида пройдет по его наибольшему диаметру, вторая — по наибольшему диаметру в плоскости, рассекающей эллипсоид посередине и перпендикулярно первой оси, и так далее. Количество осей этого эллипсоида будет равно количеству переменных, и в направлении каждой последующей оси будет все меньший и меньший разброс наблюдений. При этом количество компонент, которые исследователь выбирает как «главные», определяется произвольно. Таким образом, анализ главных компонент решает задачу сокращения количества переменных при условии сохранения максимальной доли дисперсии наблюдений.

Анализ главных компонент является исходной процедурой многих методов факторного анализа и может рассматриваться как их упрощенный аналог. Поэтому более подробно рассмотрим на его примере наиболее важные понятия факторного анализа.

В основе анализа главных компонент лежит математический метод нахождения собственных значений и собственных векторов корреляционной матрицы. Не останавливаясь на определениях и процедурах этого метода, отметим то, что действительно имеет существенное значение для дальнейшего понимания основ факторного анализа. В процессе компонентного анализа решается уравнение (в матричной форме):

$$R = A \cdot A', \tag{16.1}$$

где R — исходная матрица корреляций; A — матрица, каждый элемент которой a_{ik} — компонентная нагрузка переменной i (строка) по компоненте k (столбец); A' — транспонированная матрица A. Уравнение 16.1 Л. Терстоун назвал «фундаментальной факторной теоремой» (Г. Харман, 1972). Результатом решения этого уравнения является матрица компонентных нагрузок A.

Рассмотрим важные особенности матрицы компонентных нагрузок на примере компонентного анализа корреляционной матрицы, представленной в табл. 16.1. Решение уравнения 16.1 позволяет получить матрицу компонентных нагрузок (табл. 16.3).

Таблица 16.3

Компоненты корреляционной матрицы показателей интеллекта

Переменная	Компоненты				
	1	2	3	4	5
1	0,77	–0,58	0,00	0,03	–0,26
2	0,75	–0,60	–0,13	0,00	0,25
3	0,75	0,41	–0,06	–0,51	–0,01
4	0,68	0,53	–0,39	0,33	–0,02
5	0,78	0,30	0,52	0,18	0,05
Собственное значение (λ)	2,78	1,24	0,45	0,41	0,13
Доля дисперсии	0,56	0,25	0,09	0,08	0,02
Накопленная доля дисперсии	0,56	0,81	0,90	0,98	1,00

Собственные значения выделяются в порядке их убывания в соответствии с осями эллипсоида разброса наблюдений. Количество выделяемых компонент (и собственных значений) равно числу переменных. *Сумма* всех собственных значений равна количеству переменных. Отметим, что если бы все корреляции между исходными переменными были бы равны нулю, то каждое собственное значение равнялось бы 1. Чем выше корреляции между переменными, тем больше предыдущие собственные значения и меньше — последующие. Собственное значение, деленное на количество переменных, есть *доля дисперсии,* соответствующая данной компоненте. Все компоненты исчерпывают 100% совокупной дисперсии переменных.

Каждый элемент a_{ik} матрицы A — это **компонентная нагрузка** переменной i (строка) по компоненте k (столбец). Компонентная (как и факторная) нагрузка — аналог коэффициента корреляции, мера связи переменной i и компоненты k. Соответственно, *квадрат* компонентной нагрузки (как и корреляции) приобретает смысл части дисперсии, в данном случае — части дисперсии переменной, объясняемой соответствующей компонентой. Сумма квадратов *всех* компонентных нагрузок по строке равна 1, полной дисперсии переменной (в z-значениях).

Таким образом, полная единичная дисперсия каждой переменной разложена по компонентам. Сумма квадратов всех компонентных нагрузок *по столбцу* равна собственному значению данной компоненты:

$$\lambda_i = \sum_{j=1}^{P} a_{ij}^2, \qquad (16.2)$$

где i — номер компоненты, j — номера переменных (количеством P).

Как было указано, это собственное значение, деленное на количество переменных, есть доля дисперсии, соответствующая данной компоненте, и используется как показатель *информативности* компоненты.

Уравнение 16.1 позволяет *восстановить* коэффициенты корреляции по матрице компонентных нагрузок A, так как произведение этой матрицы на саму себя транспонированную дает корреляционную матрицу. В соответствии с правилом умножения матриц, каждый коэффициент корреляции r_{ij} может быть *восстановлен* через компонентные нагрузки, как сумма всех (по строке) произведений нагрузок для этих двух переменных по каждой компоненте. Восстановленный коэффициент корреляции вычисляется по формуле:

$$r_{ij} = \sum_{k=1}^{M} a_{ik} a_{jk}, \qquad (16.3)$$

где i, j — номера переменных в корреляционной матрице; k — номер компоненты; M — количество компонент; a — компонентные нагрузки. Так, восстановленная корреляция между переменными 3 и 5:

$$r_{35} = 0{,}75 \cdot 0{,}78 + 0{,}41 \cdot 0{,}30 + (-0{,}06)0{,}52 + (-0{,}51)0{,}18 + (-0{,}01)0{,}05 = 0{,}58.$$

Заметим, что диагональный элемент корреляционной матрицы, как корреляция признака с самим собой ($i = j$), равен сумме квадратов всех компонентных нагрузок данной переменной — по строке, то есть 1.

Исследователь может воспользоваться анализом главных компонент как упрощенным вариантом факторного анализа. Тогда он выберет не все компоненты, а только *главные*, объясняющие большую часть дисперсии. В данном случае главными будут первые две компоненты, объясняющие 81% суммарной дисперсии переменных.

Переход к главным компонентам позволяет ввести еще одно важное понятие факторного анализа. **Общность** (*Communality*) — часть дисперсии переменной, объясняемая главными компонентами (факторами), вычисляется как сумма квадратов нагрузок по строке:

$$h_i^2 = \sum_{k=1}^{M} a_{ik}^2, \tag{16.4}$$

где i — номер переменной, k — номер (главной) компоненты. Например, если по таблице 16.3 выделяются две главные компоненты, то общность переменной 1: $h_1^2 = 0{,}77^2 + (-0{,}58)^2 = 0{,}93$, а общность переменной 4: $h_4^2 = 0{,}68^2 + 0{,}53^2 = 0{,}74$. То есть первые две компоненты исчерпывают 93% дисперсии переменой 1 и 74% дисперсии переменной 4.

Восстановленные только по главным компонентам коэффициенты корреляции (по формуле 16.3) *будут меньше исходных по абсолютной величине, а на диагонали восстановленной корреляционной матрицы будут не 1, а величины общностей.*

Анализ главных компонент в «чистом виде» используется для решения одной из ключевых проблем факторного анализа — проблемы числа факторов.

Принцип выделения «главных факторов» в факторном анализе тот же, что и при анализе главных компонент. Но в отличие от компонентного анализа факторный анализ направлен на объяснение *корреляций* между переменными, а не только компонент дисперсии.

Факторная структура (*Factor Structure Matrix*) — основной результат применения факторного анализа. Элементы факторной структуры — **факторные нагрузки** (*Factor Loadings*) переменных a_{ik}, аналогичные компонентным нагрузкам (см. табл. 16.3). Однако основное требование их получения, в отличие от анализа главных компонент, — *максимально полное отражение исходных коэффициентов корреляции.* Поэтому **осноннoе уравнение факторного анализа**:

$$\hat{R} = A \cdot A' \text{ при условии } \hat{R} \to R, \tag{16.5}$$

где R — исходная матрица интеркорреляций; $\hat{R}$ — матрица *восстановленных коэффициентов корреляции*; A — матрица факторных нагрузок размерностью, столбцы которой — факторные нагрузки P переменных по M факторам; A' — транспонированная матрица A. Отличие уравнения 16.5 от сходного с ним уравнения компонентного анализа (16.1) в том, что матрица факторных нагрузок A вычисляется таким образом, чтобы *восстановленные коэффициенты корреляции минимально отличались от исходных корреляций.*

Рассмотрим искомую факторную структуру в общем виде, как матрицу факторных нагрузок (табл. 16.4). В этой таблице P строк, соответствующих переменным, и M столбцов — факторов. Значение a_{ik} — это факторная на-

грузка переменной i по фактору k. Соотношения величин в этой таблице идентично соотношениям в таблице компонентных нагрузок. *Собственное значение (Eigenvalue)* каждого фактора λ_k, по формуле 16.2, равно сумме квадратов факторных нагрузок всех переменных по фактору k (по столбцу). *Общность* каждой переменной h_i^2, в соответствии с формулой 16.4, равна сумме квадратов факторных нагрузок переменной i по всем факторам. *Коэффициент корреляции* между любыми двумя переменными может быть восстановлен по этой таблице, как сумма произведений факторных нагрузок по соответствующим строкам (по формуле 16.3).

Таблица 16.4

Факторная структура в общем виде
(M — число факторов, P — число переменных)

	1	2	...	k	...	M	h^2
1	a_{11}	a_{12}	...	a_{1k}	...	a_{1M}	h_1^2
2	a_{21}	a_{22}	...	a_{2k}	...	a_{2M}	h_2^2
...	...	...	...	...	...	...	...
i	a_{i1}	a_{i2}	...	a_{ik}	...	a_{iM}	h_i^2
...	...	...	...	...	...	...	...
P	a_{P1}	a_{P2}	...	a_{Pk}	...	a_{PM}	h_P^2
λ	λ_1	λ_2	...	λ_k	...	λ_M	...

Алгоритм факторного анализа обеспечивает максимально возможное *приближение* вычисленных, или восстановленных, коэффициентов корреляции к исходным корреляциям. Это достигается варьированием числа факторов и диагональными элементами корреляционной матрицы, на которых располагаются не единицы, как в компонентном анализе, а значения *общностей.*

Проблема числа факторов

Это первая проблема при проведении факторного анализа. Обычно заранее не известно, сколько факторов необходимо и достаточно для представления данного набора переменных. Сама же процедура факторного анализа предполагает предварительное задание числа факторов. Поэтому исследователь должен заранее определить или оценить их возможное количество. Для этого на первом этапе факторного анализа обычно применяют анализ главных компонент и используют график собственных значений (Scree plot).

На рис. 16.2 представлен график собственных значений для компонент из табл. 16.3. Компонентные нагрузки интерпретации на данном этапе не подлежат, нас интересуют только величины собственных значений.

Для определения числа факторов были предложены два критерия. Первый — **критерий Кайзера**: *число факторов равно числу компонент, собственные значения которых больше 1.*

Второй способ определения числа факторов — **критерий отсеивания Р. Кеттелла** (scree–test), требует построения графика собственных значений

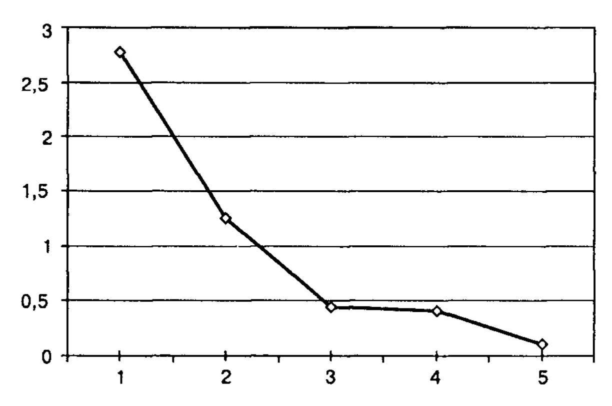


Рис. 16.2. График собственных значений для пяти показателей интеллекта

(компьютерные программы предлагают этот график при выборе метода главных компонент — scree plot). Количество факторов определяется приблизительно *по точке перегиба на графике собственных значений* до его выхода на пологую прямую после резкого спада. При этом проверяются три гипотезы: если K — точка перегиба, то возможное количество факторов равно $K-1$, K и $K+1$.

По первому критерию (Кайзера) в нашем примере число факторов равно двум, так как первые два собственных значения больше 1. По второму критерию (Р. Кеттелла) — от двух до четырех, так как точке перегиба соответствует третья компонента (см. рис. 16.2).

При определении числа факторов на практике следует помнить, что указанные критерии являются лишь примерным ориентиром. *Окончательное решение о числе факторов принимается только после интерпретации факторов.*

Проблема общности

Это вторая главная проблема факторного анализа. Единичная дисперсия каждой переменной представлена в факторном анализе как сумма ее общности и характерности:

$$1 = h_i^2 + e_i^2,$$

где h_i^2 — общность переменной с номером i; e_i^2 — ее характерность.

Общность — это часть дисперсии переменной, обусловленная действием общих факторов. **Характерность** — часть ее дисперсии, обусловленная спецификой данной переменной и ошибками измерения. Иначе говоря, общность — это полный вклад всех факторов в единичную дисперсию переменной, а характерность — это разность полной единичной дисперсии переменной и ее общности. Общность переменной i равна сумме квадратов ее нагрузок по всем M факторам (по строке факторных нагрузок):

$$h_i^2 = \sum_{k=1}^{M} a_{ik}^2.$$

Полнота факторизации — важное понятие факторного анализа, вытекающее из определения общности. Любой элемент факторной структуры — факторная нагрузка переменной, возведенная в квадрат, — приобретает смысл доли дисперсии переменной, обусловленной данным фактором. Суммируя эти доли по строке, мы получаем общность — долю дисперсии переменной, обусловленную влиянием всех общих факторов.

Суммарная дисперсия всех переменных есть сумма единичных дисперсий всех признаков, что равно просто количеству признаков. Суммируя доли дисперсии всех переменных по одному фактору, мы получаем суммарную дисперсию всех переменных, обусловленную действием этого фактора. Разделив суммарную дисперсию, обусловленную действием данного фактора, на количество признаков, мы получим долю дисперсии, обусловленную данным фактором, или **информативность** (мощность) **фактора**. Сумма квадратов всех элементов факторной структуры — факторных нагрузок — равна сумме всех общностей и суммарной дисперсии всех переменных, обусловленной общими факторами. Эта величина, деленная на количество признаков, известна как **полнота факторизации**:

$$V=\sum_{k=1}^{M}V_k=\frac{1}{P}\sum_{k=1}^{M}\lambda_k=\frac{1}{P}\sum_{i=1}^{P}h_i^2=\frac{1}{P}\sum_{k=1}^{M}\sum_{i=1}^{P}a_{ik}^2,$$

где V_k — мощность фактора с номером k; λ_k — собственное число фактора с номером i; h_i^2 — общность переменной i; a_{ik}^2 — вклад фактора i в переменную k; M — число факторов; P — число переменных.

Понятно, что качество факторного анализа тем выше, чем выше полнота факторизации. И эта величина является одним из важных показателей при выборе пользователем варианта решения, наряду с показателем того, насколько полно воспроизводятся коэффициенты корреляции. Надо отметить, что четких статистических критериев полноты факторизации не существует. Тем не менее, низкие ее значения, например меньше 0,7, свидетельствуют о желательности сокращения количества признаков или увеличения количества факторов.

Вообще говоря, **проблема общностей** заключается в том, что они, как и число общих факторов, не известны до начала анализа, но должны каким-то образом задаваться, так как величины факторных нагрузок зависят от величин общностей. Отметим, что в компонентном анализе этой проблемы не существует: общность каждой переменной равна 1, при условии выделения всех P компонент. ***Различия в методах факторного анализа и определяются тем, как решается проблема общностей.***

Методы факторного анализа

Методы факторного анализа — это различные способы получения факторной структуры при заданном числе факторов. Эти способы, как уже говорилось, отличаются решением проблемы общностей. Рассмотрим наиболее

часто применяемые методы: анализ главных компонент, метод главных факторов, факторный анализ образов (общности равны квадрату КМК), метод не взвешенных наименьших квадратов, обобщенный метод наименьших квадратов и метод максимального правдоподобия.

Анализ главных компонент (*Principal Components*) иногда используется в качестве факторного анализа, хотя это и не вполне корректно. При использовании этого метода общность каждой переменной получается автоматически, путем суммирования квадратов ее нагрузок по всем главным компонентам. Вопрос о приближении восстановленных коэффициентов корреляции к исходным корреляциям не решается. В результате факторная структура искажается в сторону преувеличения абсолютных величин факторных нагрузок.

Факторный анализ образов (*общности равны квадрату КМК*) (*Image Factoring*) — это метод главных компонент, применяемый к так называемой редуцированной корреляционной матрице, у которой вместо единиц на главной диагонали располагаются оценки общностей. Общность каждой переменной оценивается предварительно, как квадрат коэффициента множественной корреляции (КМК) этой переменной со всеми остальными. Такая оценка, с точки зрения теоретиков факторного анализа, приводит к более точным результатам, чем в анализе главных компонент. Но значения общностей недооцениваются, что также приводит к искажениям факторной структуры, хотя и меньшим, чем в предыдущем случае.

Метод главных осей (*Principal Axis Factoring*), позволяет получить более точное решение. На первом шаге общности вычисляются по методу главных компонент. На каждом последующем шаге собственные значения и факторные нагрузки вычисляются исходя из предыдущих значений общностей. Окончательное решение получается при выполнении заданного числа итераций или достижении минимальных различий между общностями на данном и предыдущем шагах.

Метод не взвешенных наименьших квадратов (*Unweighted least squares*) — минимизирует квадраты остатков (разностей) исходной и воспроизведенной корреляционных матриц (вне главной диагонали). На первом шаге оцениваются общности через квадрат КМК. Затем вычисляется факторная структура и восстанавливаются коэффициенты корреляции. Проверяется разность квадратов исходных и вычисленных корреляций. За новые значения общностей принимаются вычисленные по полученной факторной структуре. На втором шаге вычисляется новая факторная структура, и снова проверяется соответствие исходных и восстановленных коэффициентов корреляции. Процесс повторяется многократно до тех пор, пока не достигается минимально возможная разница между исходными и вычисленными корреляциями при заданном числе факторов. Метод, по определению, дает минимальные ошибки факторной структуры при фиксированном числе факторов. Реализация метода в компьютерных программах позволяет проверить расхождения между исходными и вычисленными корреляциями. Наличие многочисленных расхождений может служить дополнительным аргументом в пользу увеличения числа факторов.

Обобщенный метод наименьших квадратов (*Generalized least squares*) — отличается от предыдущего тем, что для каждой переменной вводятся специальные весовые коэффициенты. Чем больше общность переменной, тем в большей степени она влияет на факторную структуру (имеет больший вес). Это соответствует основному принципу статистического оценивания, по которому менее точные наблюдения учитываются в меньшей степени. В этом — основное преимущество этого метода перед остальными.

Метод максимального правдоподобия (*Maximum likelihood*) также направлен на уменьшение разности исходных и вычисленных корреляций между признаками. Дополнительно этот метод позволяет получить важный показатель полноты факторизации — статистическую оценку «качества подгонки». Мерой качества является оценка различия исходных и вычисленных коэффициентов корреляции по χ^2-критерию, значимость которого определяется в зависимости от числа факторов и количества переменных. Если критерий показывает значимое отклонение при M-факторной модели, следует перейти к модели с M+1 факторами, и так до тех пор, пока отклонение исходных и вычисленных корреляций перестанет быть статистически значимым по χ^2-критерию. Таким образом, χ^2-критерий позволяет определить минимально допустимое количество факторов для данного числа переменных. Однако следует помнить, что этот критерий, как и остальные формальные критерии, является дополнительным. Окончательное же решение о числе факторов принимается после содержательной интерпретации факторной структуры.

Вряд ли возможно дать общие рекомендации о преимуществе или недостатке того или иного метода. Можно лишь отметить, что анализ главных компонент дает наиболее грубое решение, а метод максимального правдоподобия позволяет статистически оценить минимально возможное число факторов для данного набора переменных. По-видимому, в каждом конкретном случае стоит сравнивать результаты применения разных методов и выбирать тот, который позволяет получить наиболее простую и доступную интерпретации факторную структуру.

Проблема вращения и интерпретации

Это третья основная проблема факторного анализа, решение которой связано с геометрическим представлением факторной структуры. Необходимость решения этой проблемы обусловлена тем, что, как правило, ***результаты факторизации непосредственно не подлежат интерпретации.*** В то же время ценность результата факторного анализа определяется прежде всего возможностью его однозначной интерпретации.

Рассмотрим результат применения метода главных осей к данным о пяти показателях способностей. Из табл. 16.5 видно, что все переменные имеют наибольшие нагрузки по первому фактору, и невозможно определить, какие переменные идентифицируют второй фактор. То есть данная факторная структура не поддается интерпретации. Для ответа на вопрос о распределении пе-

Таблица 16.5

Факторная структура пяти показателей способностей (метод главных осей, до вращения)

Переменные	Факторные нагрузки a_{ik}		Общность h_i^2
	F_1	F_2	
1	0,807	−0,482	0,88
2	0,774	−0,481	0,83
3	0,661	0,416	0,61
4	0,580	0,470	0,56
5	0,675	0,317	0,56
Собственное значение λ_k	2,478	0,959	3,44
Доля дисперсии	0,496	0,192	0,69

ременных по факторам и необходимо решить проблему вращения факторов относительно признаков.

Факторную структуру графически можно представить в виде точек-признаков в пространстве *M* факторов. Положение каждой точки задается факторными нагрузками как координатами этой точки по соответствующим осям-факторам. Для нашего примера такое графическое изображение факторной структуры представлено на рис. 16.3.

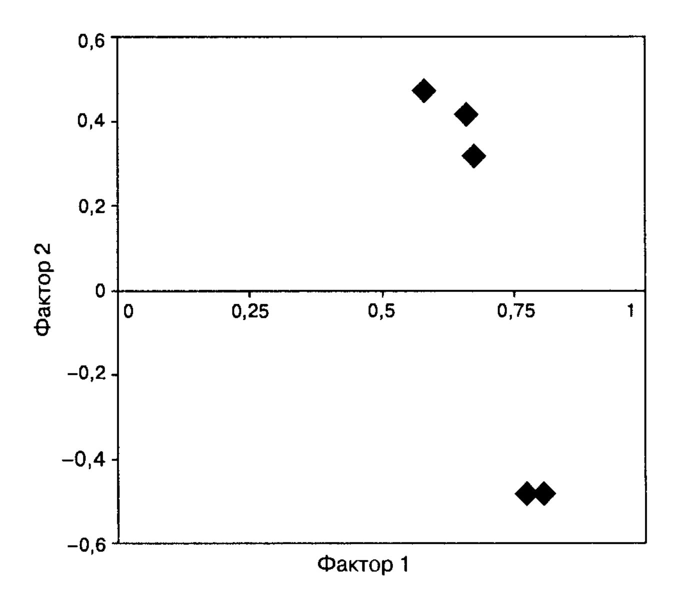
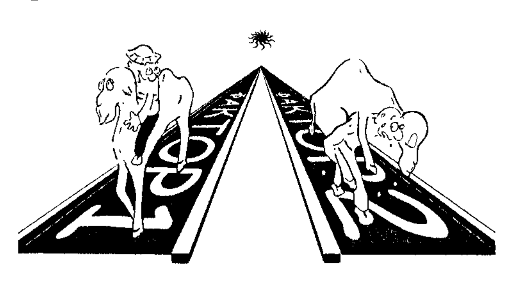


Рис. 16.3. График пяти показателей интеллекта в осях двух факторов до вращения

Расстояние каждой точки от начала координат или длина *вектора-переменной* равны сумме квадратов всех координат этой точки (конца вектора-переменной). Поскольку координаты — это факторные нагрузки, то длина каждого вектора равна общности соответствующей переменной.

Без доказательства укажем, что коэффициент корреляции между каждой парой переменных равен *косинусу угла* между соответствующими векторами в пространстве общих факторов. Иначе говоря, чем выше корреляция, тем меньше угол между соответствующими переменными.

Указанные соотношения между переменными в осях факторов никак не изменятся, если мы повернем оси факторов на любой угол относительно переменных, при условии соблюдения взаимной ортогональности (перпендикулярности) факторов. Из этого следует вывод, что мы можем поворачивать факторы относительно переменных как угодно, соблюдая ортогональность факторов. При этом наиболее предпочтительно, чтобы каждая переменная в результате вращения оказалась вблизи оси фактора, иными словами, имела бы максимальную нагрузку по одному фактору и минимальные — по всем остальным. Только в этом случае каждая переменная будет соотнесена только с одним фактором, что и требуется для интерпретации факторной структуры.



Каждая переменная имеет большую нагрузку только по одному из факторов

В нашем примере (рис. 16.3) желателен поворот осей факторов по часовой стрелке так, чтобы фактор 1 прошел вблизи переменных 1 и 2, а фактор 2 — вблизи переменных 3–5. *Решение, при котором каждая переменная имеет большую нагрузку только по одному фактору, а по остальным ее нагрузки близки к нулю, называется* ***простой структурой.***

На заре появления многофакторного анализа проблема вращения решалась графически. Чертились графики факторной структуры — по одному для каждой пары факторов. Затем делали графический поворот осей факторов относительно переменных, после чего линейкой измеряли новые проекции переменных на эти оси. Таким образом получали факторную структуру после вращения.

В настоящее время используются аналитические способы вращения, реализованные во всех компьютерных программах факторного анализа. Работа аналитических методов подобна геометрическому вращению «вручную». Каждая пара факторов поворачивается относительно переменных до тех пор, пока не достигается наиболее возможная простота структуры. В одних случаях критерием простоты является факторная сложность переменных (**квартимакс**), в других — индекс сложности каждого фактора (**варимакс**), где факторная сложность переменной пропорциональна числу общих факторов, связанных с ней, а индекс сложности фактора пропорционален числу переменных, с ним связанных. Наиболее широко применяется вращение, где на каждом шаге простота структуры определяется по критерию варимакс Г. Кайзера — **варимакс-вращение**.

Результат применения варимакс-вращения к факторной структуре из табл. 16.5 представлен в табл. 16.6; графический результат — на рис. 16.4.

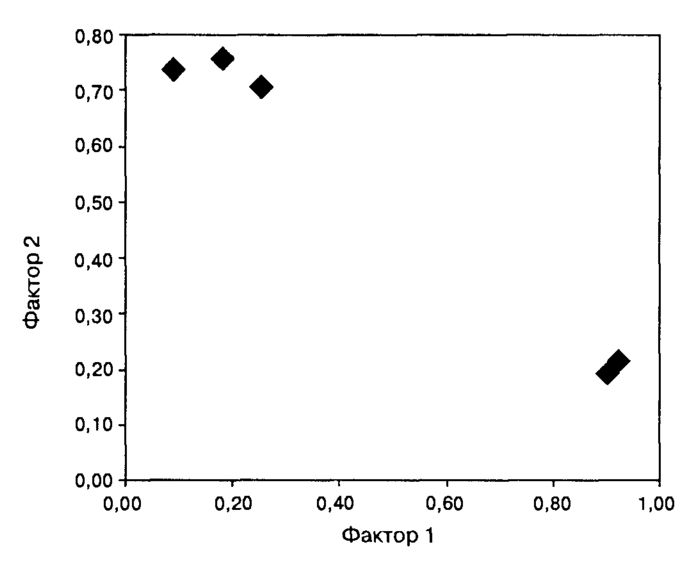


Рис. 16.4. Факторная структура пяти показателей интеллекта после варимакс-вращения

После вращения каждая переменная имеет большую нагрузку только по одному фактору. Следовательно, каждый фактор может быть однозначно интерпретирован через входящие в него переменные: фактор 1 — по признакам 1 и 2, фактор 2 — по признакам 3 и 5. Так как переменная 1 — счет в уме, переменная 2 — числовые ряды, то фактор 1 может быть идентифицирован как «арифметические способности». Переменные, входящие в фактор 2 (осведомленность, словарный запас, сходство), позволяют интерпретировать его как фактор словесного понимания.

Таблица 16.6

Факторная структура пяти показателей интеллекта после варимакс-вращения

Переменные	F_1	F_2	h^2
1	0,921	0,217	0,88
2	0,903	0,193	0,83
3	0,183	0,757	0,61
4	0,088	0,739	0,56
5	0,260	0,700	0,56
СКН*	1,773	1,694	3,44
Доля дисперсии	0,355	0,339	0,69

* СКН — сумма квадратов факторных нагрузок [1]

[1] После вращения сумма квадратов факторных нагрузок по столбцу не равна собственному значению фактора.

Проблема оценки значений факторов

После интерпретации факторной структуры допустима **оценка значений факторов** для объектов. Это позволяет перейти от множества исходных переменных к существенно меньшему числу факторов как новых переменных. Это может понадобиться исследователю как для более компактного представления различий между объектами (или их группами), так и для дальнейшего анализа — регрессионного, дисперсионного и т. д. В этом смысле факторный анализ как общенаучный метод выполняет задачу сокращения размерности набора переменных с минимальными потерями исходной информации.

В качестве *оценки значения фактора i* для объекта k используется линейная комбинация значений исходных переменных X:

$$f_{ik} = \sum_{j=1}^{P} \beta_{i,j} x_{jk} = \beta_{1i} x_{1k} + \beta_{2i} x_{2k} + \ldots + \beta_{Pi} x_{Pk}, \qquad (16.6)$$

где f_{ik} — значение фактора с номером i для объекта k; x_{jk} — значение признака с номером j для этого объекта; $\beta_{i,j}$ — факторный коэффициент признака j для фактора i.

Поскольку известны корреляции между исходными переменными и корреляции этих переменных с факторами (факторные нагрузки), то в качестве наиболее состоятельной **оценки факторных коэффициентов** $\beta_{i,j}$ чаще всего выступают *коэффициенты множественной регрессии*. В качестве «зависимой» переменной выступает фактор, в качестве «независимых» — исходные переменные, а вычисления производятся по формуле 15.3.

Вычисленные по модели множественной регрессии оценки факторных коэффициентов далее используются для вычисления всех оценок значений факторов для каждого объекта по формуле 16.6. Таким образом, исследователь переходит от множества P переменных к небольшому числу M новых переменных-факторов, интерпретируемых через исходные переменные. Отметим, что средние значения каждой такой новой переменной для всех объектов равно 0, а стандартное отклонение близко (но меньше) 1.

Проблема оценки значений факторов связана с тем, что невозможно точно выразить общий фактор через исходные переменные, так как каждая из этих переменных содержит помимо общности и характерную часть, которую невозможно отделить. Поэтому можно получить лишь *оценку* значений факторов по исходным переменным, *надежность* которой обладает большей или меньшей определенностью — в зависимости от вида исходных данных и факторной структуры.

Зависимость надежности факторных оценок от факторной структуры выражается в следующем. Чем меньше суммарная общность всех переменных, тем меньше надежность факторных оценок всех факторов. Чем меньше информативность фактора (сумма квадратов факторных нагрузок по столбцу), тем меньше надежность факторных оценок данного фактора.

В связи с надежностью факторных оценок особое значение приобретает качество измерения исходных переменных. *Чем больше исходные переменные*

соответствуют требованиям, которые предъявляются к метрическим переменным, тем надежнее факторные оценки. Если переменные измерены в порядковой или, тем более, в дихотомической шкале, то надежность факторных оценок снижается до непредсказуемого уровня. Тем не менее, некоторые авторы (см. К. Иберла, 1980) на основе факторного моделирования доказывают, что в случае исходных порядковых и даже дихотомических данных факторные оценки могут быть состоятельными. *Условиями состоятельности* факторных оценок являются действительно простая факторная структура, а также высокие значения общностей и факторных нагрузок переменных.

В заключение обзора математико-статистических идей факторного анализа заметим, что современный факторный анализ — изящная математическая процедура, имеющая достаточное статистическое обоснование. Это выгодно отличает данный метод от остальных. Эта изящность, наряду с исходным психологическим обоснованием, однако, часто вводит в заблуждение новичка, ожидающего получить «готовый ответ» в результате применения факторного анализа. Необходимо помнить, что факторный анализ не добавляет никакой новой информации к эмпирическим данным. Его задача — в обеспечении *возможности* интерпретировать данные. Качество же интерпретации целиком зависит от исследователя, от того, насколько и как он понимает исходные измерения, основы и процедуру факторного анализа.

ПОСЛЕДОВАТЕЛЬНОСТЬ ФАКТОРНОГО АНАЛИЗА

Особенность факторного анализа заключается в неопределенности решения его основных проблем, изложенных в предыдущем параграфе. Нет четких критериев качества их решения, есть лишь рекомендации, которыми руководствуется исследователь в своем стремлении содержательно интерпретировать получаемые результаты. Поэтому факторный анализ — это пошаговая процедура, где на каждом шаге исследователь принимает решение о дальнейших преобразованиях данных. Главным же ориентиром на этом пути остается возможность получения содержательной интерпретации конечных результатов.

Весь процесс факторного анализа можно представить как выполнение шести этапов:

1. Выбор исходных данных.
2. Предварительное решение проблемы числа факторов.
3. Факторизация матрицы интеркорреляций.
4. Вращение факторов и их предварительная интерпретация.
5. Принятие решения о качестве факторной структуры.
6. Вычисление факторных коэффициентов и оценок.

Исследователь, в зависимости от своих целей, решает, сколько раз повторить эту последовательность, какие из этапов будут пропущены и насколько глубоко будет проработан каждый из них. Например, если исследователя интересует только структура взаимосвязей признаков, то достаточно выполнить эту последовательность один раз, без последнего этапа.

Этап 1. Выбор исходных данных

Модель факторного анализа разрабатывалась для метрических данных. Поэтому первое требование к исходным данным — представление всех признаков в метрической шкале (не обязательно с одинаковыми средними и дисперсиями).

Включение в анализ порядковых или бинарных данных допустимо, но исследователь должен отдавать себе отчет, что искажения факторной структуры при этом будут соответствовать искажениям коэффициентов корреляций, и характер этих искажений неизвестен. В общем случае желательно перейти к единой шкале для всех признаков, либо ранговой, либо бинарной, затем вычислять матрицу интеркорреляций, выбирая соответствующие меры взаимосвязи. Исследователь потеряет при этом существенную долю исходной информации. А о допустимости дальнейшего вычисления значений факторных коэффициентов и оценок для объектов известно мало. Можно лишь сказать, что достоверность и ценность конечного результата обратно пропорциональны доле потерянной исходной информации.

Если цель факторного анализа заключается только в определении структуры взаимосвязей переменных, то допустимо применение порядковых данных, но перед проведением факторного анализа необходимо перейти к рангам по каждой переменной. Допустимо также использовать факторный анализ в отношении дихотомических переменных, если задача ограничивается определением структуры взаимосвязей и дихотомические корреляции между переменными не очень велики (не превышают 0,7)[1].

Порядковые и даже дихотомические данные могут использоваться для вычисления оценок факторов, но при условии действительно простой факторной структуры, высоких значениях общностей и факторных нагрузок переменных, определяющих каждый фактор (К. Иберла, 1980). При этом желательно проверять устойчивость факторной структуры на параллельных выборках.

Как и в других многомерных методах, недопустимы функциональные зависимости между переменными и корреляции, близкие к единице.

Количественное соотношение признаков и объектов зависит от целей исследования. Если цель анализа — изучение структуры взаимосвязей признаков, уменьшение их исходного количества путем перехода к новым пере-

[1] См.: Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. М., 1989. С. 64–65.

менным — факторам, то строгих ограничений нет. Желательно лишь, чтобы количество признаков было не меньше количества объектов. Если исследователь хочет обнаружить и обосновать наличие факторов за взаимосвязями переменных, то желательно иметь в три раза больше объектов, чем признаков. Данное соотношение может сложиться и в процессе анализа — при отсеивании мало информативных переменных. Если же стоит задача обоснования выявленной факторной структуры для генеральной совокупности, то объектов должно быть еще больше, для проверки устойчивости этой структуры на параллельных выборках.

Этап 2. Решение проблемы числа факторов

На этом этапе матрица интеркорреляций исходных признаков обрабатывается с использованием анализа главных компонент. Применяется критерий отсеивания Р. Кеттелла и критерий Кайзера — величины собственного значения фактора, большего 1 (Eigenvalue, $\lambda > 1$). Эти критерии не являются жесткими, поэтому далее проверяется несколько гипотез о числе факторов. Начинать при этом рекомендуется с максимально возможного числа факторов, с учетом обоих критериев, постепенно уменьшая их число.

Этап 3. Факторизация матрицы интеркорреляций

Выбирается метод факторизации, желательно — главных осей, наименьших квадратов или максимального правдоподобия. Задается число факторов, в соответствии с проверяемой гипотезой. Результатом данного этапа является матрица факторных нагрузок (факторная структура) до вращения, которая не подлежит интерпретации.

Полезной информацией на этом этапе могут являться суммарная доля дисперсии (информативность) факторов и значения общностей переменных. Суммарная доля дисперсии — показатель того, насколько полно выделяемые факторы могут представить данный набор признаков, а этот набор — выделяемые факторы. Общность переменной — показатель ее «участия» в факторном анализе, насколько она влияет на факторную структуру. Переменные с наименьшими общностями — ближайшие кандидаты на исключение из анализа в дальнейшем.

Этап 4. Вращение факторов и их предварительная интерпретация

На этом этапе выбирается один из аналитических методов вращения факторов, обычно — варимакс-вращение (Varimax normalized). Существуют и другие методы вращения, в том числе косоугольного, но они выходят за рамки

нашего рассмотрения. В результате вращения достигается факторная структура, наиболее доступная для интерпретации при данном соотношении переменных и факторов.

Интерпретация факторов производится по таблице факторных нагрузок после вращения в следующем порядке. *По каждой переменной* (строке) выделяется наибольшая по абсолютной величине нагрузка — как доминирующая. Если вторая по величине нагрузка в строке отличается от уже выделенной менее чем на 0,2, то и она выделяется, но как второстепенная. После просмотра всех строк — переменных, начинают просмотр столбцов — факторов. *По каждому фактору* выписывают наименования (обозначения) переменных, имеющих наибольшие нагрузки по этому фактору — выделенных на предыдущем шаге. При этом обязательно учитывается знак факторной нагрузки переменной. Если знак отрицательный, это отмечается как противоположный полюс переменной. После такого просмотра всех факторов каждому из них *присваивается наименование*, обобщающее по смыслу включенные в него переменные. Если трудно подобрать термин из соответствующей теории, допускается наименование фактора по имени переменной, имеющей по сравнению с другими наибольшую нагрузку по этому фактору.

Этап 5. Принятие решения о качестве факторной структуры

Формальное требование к факторной структуре сформулировал Л. Терстоун еще в 1930-х годах, назвав его принципом простой структуры. Геометрически **принцип простой структуры** означает, что все переменные располагаются на осях факторов, то есть каждая переменная имеет близкие к нулю нагрузки по всем факторам, кроме одного. На практике достижение такого результата с первого раза маловероятно, но качество факторной структуры определяется степенью приближения к простой структуре.

Следует отметить общий принцип соотношения качества факторной структуры и качества исходных данных: чем ниже качество исходных данных в смысле требований, предъявляемых к метрическим переменным, тем выше требования к простоте факторной структуры, величине общностей и факторных нагрузок.

В настоящее время не существует формальных критериев соответствия факторной структуры простой. Поэтому основным критерием остается возможность хорошей содержательной интерпретации каждого фактора по двум и более исходным переменным. Если перед исследователем стоит дополнительно проблема обоснования устойчивости (воспроизводимости) факторной структуры в генеральной совокупности, то добавляется требование однозначного соотнесения каждой переменной с одним из факторов. Это требование означает, что каждая переменная имеет большую по абсолютной величине нагрузку (0,7 и выше) только по одному фактору и малые (0,2 и менее) — по всем остальным.

Можно предложить способы максимального приближения к простой структуре путем пошагового сокращения числа факторов и переменных.

1. Если по результатам интерпретации выявлен фактор, по которому ни одна из переменных не получила максимальной нагрузки (по строке), то это свидетельствует о необходимости сокращения количества факторов на один и повторения этапов 3 и 4 с новым числом факторов. То же касается фактора, идентифицируемого лишь по одной переменной, когда остальные в него не попадают даже с второстепенными нагрузками.

2. Определяются неоднозначные переменные. Каждая такая переменная имеет примерно одинаковые по абсолютной величине максимальные нагрузки по двум и более факторам. Если обосновывается устойчивость факторной структуры, то неоднозначной является переменная, у которой между максимальной и следующей за ней по величине нагрузкой разность менее 0,5. Неоднозначные переменные поочередно удаляются из числа исходных переменных, и каждый раз повторяются этапы 3 и 4.

Очевидно, что приближение к простой структуре связано с невосполнимой потерей исходной эмпирической информации. И каждый раз исследователь должен решать, насколько целесообразна эта потеря в свете стоящих перед ним задач. Наиболее жестки требования к простой структуре в случае обоснования устойчивости и воспроизводимости факторов, например, при разработке теста или факторной теоретической модели. Гораздо мягче требования при решении наиболее часто встречающихся задач — при изучении структуры взаимосвязей или при сокращении исходного набора признаков для дальнейшего исследования, например, различий между группами объектов.

Этап 6. Вычисление факторных коэффициентов и оценок

Это заключительный, наиболее однозначный и простой этап факторного анализа.

Оценки факторных коэффициентов являются коэффициентами линейного уравнения, связывающего значение фактора и значения исходных переменных. Они показывают, с каким весом входят исходные значения каждой переменной в оценку фактора. Факторные коэффициенты можно использовать для вычисления факторных оценок для новых объектов, не включенных ранее в факторный анализ.

Факторные оценки — значения факторов для каждого объекта (испытуемого). Их получение чаще всего и является конечным результатом факторного анализа. Вычисленные оценки факторов, как новые переменные, являются независимыми, отражающими реальную структуру взаимосвязей исходных признаков и наиболее полно передающими исходную эмпирическую информацию. В этом факторные оценки выгодно отличаются от других способов интегрирования исходных данных, например от простого суммирования пунктов теста или анкеты в шкалы.

ПРИМЕР 16.2

До широкого распространения персональных компьютеров полновесный факторный анализ был экзотической, весьма трудоемкой многоэтапной процедурой, когда очередной шаг исследователь выбирает по результатам выполнения предыдущих этапов. В настоящее время можно контролировать процесс факторного анализа без «посредников» (программиста, операторов), пользуясь современным программным обеспечением. Для этого не нужны знания программиста и математика, достаточны осведомленность в основных математико-статистических идеях метода и умение «читать» промежуточные и конечные результаты факторного анализа. При этом факторный анализ может быть рекомендован для решения очень широкого круга не только исследовательских, но и практических задач. Перечислим некоторые из них.

1. Факторный анализ как инструмент интерпретации позволяет быстро выделить группировки (кластеры) взаимосвязанных переменных, решая проблемы корреляционного анализа: наличия множества переменных и множества статистических проверок.

2. Факторный анализ как альтернатива простого суммирования значений исходных переменных позволяет учитывать реальную структуру данных и избегать излишних потерь драгоценной исходной информации. Затраты времени и сил на такую обработку данных при помощи факторного анализа часто меньше, чем при суммировании баллов «вручную». При этом выигрыш весьма ощутим — в детальности и корректности получаемых результатов.

3. Как подготовительный этап для прогнозирования факторный анализ позволяет получить некоррелированные интегральные переменные (факторы), наиболее пригодные для применения в регрессионном или дискриминантном анализе.

4. При исследовании индивидуальных или межгрупповых различий по множеству признаков факторный анализ позволяет сократить исходное множество признаков до нескольких факторов, по которым различия проявляются наиболее ярко.

Одно из направлений психодиагностики, в котором факторный анализ является незаменимым инструментом, — это исследования с применением семантического дифференциала в его многочисленных модификациях. Как известно, само появление этого метода основано на модели факторного анализа, и его применение в психосемантических исследованиях вошло в традицию.

Рассмотрим результаты факторного анализа в последовательности их получения на фрагменте реального психологического исследования с использованием метода семантического дифференциала. В исследовании применялся личностный дифференциал (ЛД), разработанный в ЛНИПИ им. В. М. Бехтерева. ЛД включает в себя 21 пару прилагательных, сгруппированных по трем шкалам: О — оценки, С — силы и А — активности. Исследование проводилось на слушателях психологического специального факультета, проходящих переподготовку по психологии (43 человека). Изучались ожидаемые слушателями изменения в восприятии себя в результате смены специальности. Соответственно им предлагалось оценить себя дважды при помощи личностного дифференциала: а) в настоящий момент; б) какими они себя видят через пять лет. Путем традиционного усреднения оценок по факторам было обнаружено, что слушатели ожидают существенных изменений только в направлении усиления активности (экстравертированности). Более детальные результаты дал факторный анализ.

На первом этапе вся выборка была разделена на две части в случайном порядке (21 и 22 человека). По результатам факторизации 21 признака дважды, для двух вы-

борок, объемом 42 (21×2) и 44 (22×2) объекта, были отобраны 10 пар прилагательных. Критерием отбора была сходная группировка этих признаков по трем факторам, то есть воспроизводимость факторной структуры. Список этих прилагательных приведен как фрагмент бланка в табл. 16.7 с указанием их принадлежности к одной из трех шкал по традиционному варианту методики.

Таблица 16.7

Отобранные пары прилагательных личностного дифференциала с «ключом» по исходному варианту методики

№	Бланк ЛД (пары прилагательных)			«Ключ»
1	Разговорчивый	3210123	Молчаливый	А
2	Безответственный	3210123	Добросовестный	О
3	Замкнутый	3210123	Открытый	А
4	Зависимый	3210123	Независимый	С
5	Деятельный	3210123	Пассивный	А
6	Вялый	3210123	Энергичный	А
7	Расслабленный	3210123	Напряженный	С
8	Суетливый	3210123	Спокойный	А
9	Несамостоятельный	3210123	Самостоятельный	С
10	Раздражительный	3210123	Невозмутимый	А

Для дальнейшего анализа использовалась объединенная выборка: 10 признаков для 86 объектов (43×2). После вычисления матрицы корреляций между отобранными признаками был проведен компонентный анализ для уточнения числа факторов. Полученный в результате график собственных значений приведен на рис. 16.5.

Выше единичного собственного значения лежат три фактора, изгиб графика наблюдается на уровне четвертого фактора. Следовательно, ожидаемое количество факторов — от 3 до 5. При проверке 4- и 5-факторных структур ни один из признаков не имел максимальные нагрузки по последним факторам, то есть не идентифицировался по переменным. Вывод — достаточное количество факторов равно трем.

В качестве метода факторизации был выбран метод максимального правдоподобия, позволяющий оценить статистическую значимость «качества подгонки» — полноты факторизации по распределению остаточных коэффициентов корреляции.

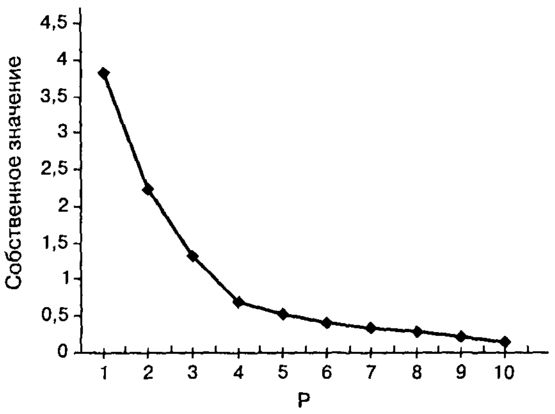


Рис. 16.5. График собственных значений после компонентного анализа 10 шкал ЛД

Действительно, при двухфакторном решении статистическая значимость составляет $p = 0{,}05$, а при трехфакторном $p = 0{,}731$. Это свидетельствует в пользу выбора трехфакторного решения. Полученная факторная структура до вращения (Unrotated) приведена в левой части табл. 16.8.

Таблица 16.8

Факторная структура 10 шкал ЛД до и после варимакс-вращения

Номер переменной	Факторные нагрузки (Factor Loadings)						Общность (h^2)
	До вращения (Unrotated)			После вращения (Varimax normalized)			
1	–0,12	0,62	0,53	0,09	–0,09	–0,82	0,68
2	0,19	–0,41	0,36	0,02	0,58	0,04	0,34
3	0,45	–0,55	–0,42	0,23	0,20	0,76	0,68
4	0,70	0,13	0,15	0,70	0,21	–0,03	0,53
5	–0,24	0,68	–0,15	0,03	–0,63	–0,39	0,55
6	0,47	–0,66	0,12	0,20	0,65	0,45	0,67
7	–0,93	–0,22	0,06	–0,94	–0,07	–0,11	0,91
8	0,73	0,36	–0,11	0,81	–0,11	0,01	0,67
9	0,49	–0,44	0,53	0,28	0,80	0,01	0,72
10	0,74	–0,06	0,11	0,66	0,32	0,13	0,56
СКН	3,19	2,15	0,96	2,66	2,01	1,63	6,3
Доля дисперсии	0,32	0,21	0,10	0,27	0,20	0,16	0,63

Суммарная информативность всех трех факторов, равная сумме собственных значений, деленной на количество переменных, составляет 0,63. Иными словами, выделенные факторы объясняют 63% суммарной дисперсии признаков — более половины, что считается приемлемым результатом.

В правой части табл. 16.8 приведена факторная структура 10 признаков после варимакс-вращения (Varimax normalized). Все признаки однозначно соотносятся по высоким факторным нагрузкам только с одним из факторов. Большинство признаков по другим факторам имеет незначительные (менее 0,2) факторные нагрузки. Можно заключить, что полученная факторная структура является достаточно простой, и приступить к интерпретации факторов. Перед интерпретацией по каждой переменной (по строке) отмечается наибольшая по абсолютной величине факторная нагрузка.

Положительный полюс фактора интерпретируется по положительным полюсам переменных, имеющих наибольшие положительные нагрузки, и по отрицательным полюсам переменных, имеющих отрицательные (наибольшие по абсолютной величине) нагрузки по этому фактору. Отрицательному полюсу фактора соответствуют отрицательные полюса переменных с положительными и положительные полюса — с отрицательными наибольшими по абсолютной величине нагрузками по этому фактору.

Фактор 1 имеет наибольший вес или наибольшую информативность (27%). Его положительный полюс определяется положительными полюсами переменных 4 (независимый), 8 (спокойный), 10 (невозмутимый) и отрицательным полюсом переменной 7 (расслабленный). Отрицательный полюс этого фактора определяется про-

тивоположными полюсами указанных переменных: зависимый, суетливый, раздражительный, напряженный. Таким образом, на положительном полюсе фактора — спокойная невозмутимость, на отрицательном — раздражительная слабость. В соответствии с известной психологической терминологией, этот фактор может быть идентифицирован как фактор эмоциональной стабильности — нейротизма.

Фактор 2 (информативность 20%): положительный полюс определяется положительными полюсами переменных 9 (самостоятельный), 6 (энергичный), 2 (добросовестный) и отрицательным полюсом переменной 5 (деятельный). Отрицательный полюс этого фактора, соответственно: несамостоятельный, вялый, безответственный, пассивный. Этот фактор можно обозначить как добросовестность (ответственность) — несамостоятельность.

Фактор 3 (информативность 16%): его положительные значения определяются положительным полюсом переменной 3 (открытый) и отрицательным — переменной 1 (разговорчивый). Этот фактор также положительно коррелирует с переменной 6 (энергичный). Отрицательный полюс фактора: замкнутый, молчаливый (вялый). Этими прилагательными обычно обозначают проявление экстраверсии (положительный полюс) — интроверсии (отрицательный полюс).

Выделенные три фактора, таким образом, однозначно интерпретируются, соотносясь с общепринятыми психологическими понятиями. Интересно их соотношение с факторами так называемой большой пятерки[1], связывающей «обыденные» и «научные» теории личности. Эти три фактора полностью соответствуют трем из пяти факторов «пятерки»: экстраверсии, нейротизму и добросовестности (ответственности). Два остальных фактора из «пятерки» (оригинальность и альтруизм) не нашли отражения, по-видимому, в связи с неполнотой списка прилагательных в ЛД.

Факторы, получившие исчерпывающую интерпретацию, далее были вычислены для каждого объекта. Для вычисления факторных оценок для объектов (испытуемых) достаточно выбрать соответствующую функцию в меню программы факторного анализа, не выходя из нее после выполнения предыдущего шага. Вычислением факторных оценок и завершился собственно факторный анализ самооценок слушателей.

Затем по вычисленным факторным оценкам были сопоставлены две выборки ответов: «Я сейчас» и «Я через 5 лет». При помощи t-критерия Стьюдента для зависимых выборок были получены статистически значимые различия между этими выборками по всем трем факторам.

Опуская психологическую интерпретацию конечного результата, отметим то главное, что позволил сделать факторный анализ. При помощи него удалось выявить действительную семантическую структуру самооценки, существенно отличающуюся от заданной заранее в ЛД. В аморфном понятии «активность», выражающем различие в восприятии настоящего и будущего «Я», факторный анализ выявляет три различные тенденции к изменениям в будущем, по которым и наблюдаются наибольшие индивидуальные различия.

---

[1] Во многих странах исследования выявили очень сходную структуру, состоящую из пяти факторов. Эти факторы достаточно полно передают основные значения прилагательных, с помощью которых люди описывают себя и других: *экстраверсия, альтруизм, добросовестность, нейротизм, оригинальность*. (См.: *Лаак Я. тер*. Психодиагностика: проблемы содержания и методов. М., 1996; *Шмелев А. Г.* Психодиагностика личностных черт. СПб., 2002.)

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Рассмотрим последовательность применения факторного анализа (программа SPSS), при помощи которого были получены результаты приведенного выше примера 16.2. В качестве исходных данных (Data Editor) были использованы 10 переменных (v1, v2, ..., v10), измеренных для 86 объектов. Интерпретация основных показателей исчерпывающе изложена при обсуждении последнего примера.

1. Выбираем **Analyze > Data Reduction > Factor...**

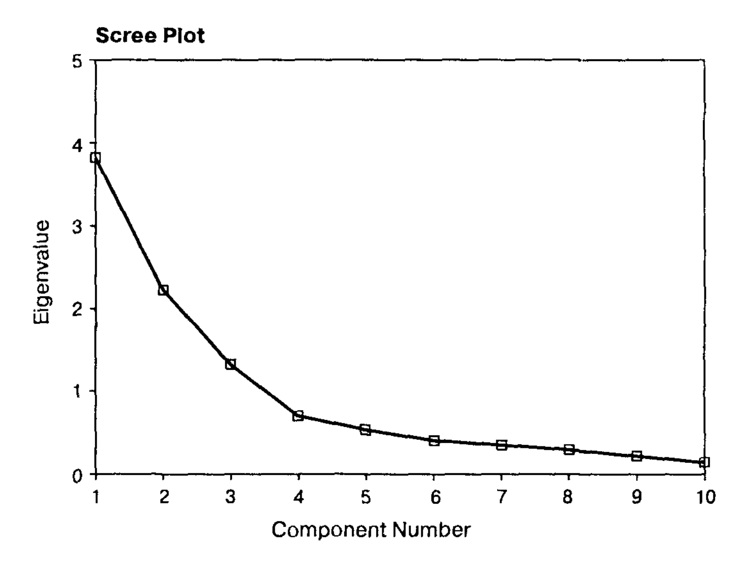
2. В открывшемся окне диалога переносим из левого в правое верхнее окно (**Variables**) *переменные*, необходимые для факторного анализа (v1, ..., v10).

3. В том же окне выбираем метод.

Примечание. В связи с тем, что на первом этапе необходимо предварительно оценить число факторов, нужно выбрать анализ главных компонент и получить график собственных значений.

Нажимаем кнопку **Extraction...** (Метод факторизации). В открывшемся окне диалога убеждаемся, что по умолчанию установлено: **Principal components** (Главные компоненты). Флажком отмечаем **Scree plot** (График собственных значений). Нажимаем **Continue** (Продолжить). Нажимаем ОК.

4. Получаем результаты, из которых на данном этапе нас интересует только *график собственных значений*:



Примечание. После предварительного определения числа факторов по этому графику начинаем факторный анализ сначала.

5. Выбираем **Analyze > Data Reduction > Factor...** В открывшемся окне диалога убеждаемся, что выбранные ранее переменные сохранены для анализа в окне **Variables** (переменные).

6. Задаем *число факторов* и выбираем *метод факторного анализа*.

Нажимаем кнопку **Extraction...** (Метод факторизации). В открывшемся окне диалога задаем необходимое число факторов: отмечаем щелчком мыши **Number of factors** (Число факторов) и в соответствующем окне указываем: 3. Выбираем метод: нажимаем кнопку **Method: ...** ▷ и выбираем **Maximum likelihood** (Максимального правдоподобия). Флажок **Scree plot** (График собственных значений) можно снять. Нажимаем **Continue** (Продолжить).

7. Задаем *метод вращения* факторов. Нажимаем кнопку **Rotation...** (Вращение) и назначаем метод вращения — варимакс: в поле **Method** (Метод) нажатием левой кнопки мыши отмечаем **Varimax** (Варимакс). Нажимаем **Continue** (Продолжить).

8. Для удобства анализа факторных нагрузок можно задать вывод только тех из них, которые больше 0,1. Для этого нажимаем кнопку **Options...** (Опции...) и в поле **Coefficient display format** (Формат выводимых коэффициентов) отмечаем левой кнопкой мыши **Suppress absolute values less then 0,1** (Скрыть абсолютные значения менее чем 0,1). Нажимаем **Continue** (Продолжить). Нажимаем ОК.

9. Получаем результаты:

А) Значения общностей (h^2):

Communalities

	Initial	Extraction
v1	.525	.684
v2	.378	.337
v3	.591	.675
v4	.543	.533
v5	.534	.547
v6	.622	.669
v7	.760	.910
v8	.634	.673
v9	.544	.716
v10	.579	.559

Extraction Method: Maximum Likelihood.

Начальные значения общностей (Initial) равны КМК; конечные значения общностей (Extraction) получены в результате факторизации методом максимального правдоподобия.

В) Компоненты суммарной дисперсии (см. табл. Total Variance Explained).

В этой таблице указаны распределенные по факторам: исходные собственные значения (Initial Eigenvalues); суммы квадратов факторных нагрузок для извлеченных факторов (Extraction Sums of Squared Loadings); суммы квадратов факторных нагрузок после вращения (Rotation Sums of Squared Loadings). Для собственных значений и сумм квадратов по каждому фактору указаны: сами значения (Total), *процент от общей дисперсии* (% of Variance) и *накоплен-*

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.815	38.155	38.155	3.191	31.914	31.914	2.659	26.591	26.591
2	2.227	22.266	60.421	2.147	21.474	53.388	2.010	20.097	46.688
3	1.320	13.196	73.617	.964	9.643	63.031	1.634	16.343	63.031
4	.692	6.924	80.541						
5	.533	5.328	85.870						
6	.402	4.016	89.885						
7	.341	3.412	93.298						
8	.295	2.955	96.253						
9	.221	2.209	98.462						
10	.154	1.538	100.000						

Extraction Method: Maximum Likelihood.

ный процент от общей дисперсии (Cumulative %). (Последние два показателя важны для определения информативности факторов.)

C) Факторные нагрузки до вращения:

Factor Matrix(a)

	Factor		
	1	2	3
v1	-.125	-.620	.533
v2	.188	.411	.364
v3	.445	.548	-.420
v4	.703	-.131	.147
v5	-.240	-.684	-.148
v6	.469	.660	.116
v7	-.926	.222	
v8	.731	-.356	-.110
v9	.486	.440	.534
v10	.737		.111

Extraction Method: Maximum Likelihood.
a 3 factors extracted. 6 iterations required.

D) Статистическая оценка «качества подгонки»:

Goodness-of-fit Test

Chi-Square	df	Sig.
13.963	18	.731

Статистическая значимость (Sig.) позволяет судить о том, достаточное ли количество факторов выделено. В данном случае p(Sig.) > 0,05, следовательно, трех факторов достаточно.

E) Факторные нагрузки после вращения:

Rotated Factor Matrix(a)

	Factor		
	1	2	3
v1			-.818
v2		.579	
v3	.232	.204	.761
v4	.699	.210	
v5		-.628	-.390
v6	.198	.654	.450
v7	-.945		-.110
v8	.813	-.110	
v9	.279	.798	
v10	.661	.325	.126

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization.
a Rotation converged in 5 iterations.
(Указаны факторные нагрузки, превышающие по абсолютной величине 0,1.)

Примечание. После получения окончательного варианта результатов факторного анализа и интерпретации факторов (по таблице факторных нагрузок после вращения) возможно вычисление факторных оценок — как новых переменных для последующего анализа другими методами.

10. Вычислим факторные оценки как новые переменные. Для этого необходимо начать факторный анализ с самого начала, сохраняя все последние установки.

A) Выбираем **Analyze > Data Reduction > Factor...** В открывшемся окне диалога убеждаемся, что выбранные ранее переменные сохранены для анализа в окне **Variables** (переменные). Нажимаем **Extraction...** и убеждаемся, что выбран метод максимального правдоподобия (**Maximum likelihood**) и задано три фактора (**Number of factors:** 3). Нажимаем **Continue**. Нажимаем **Rotation...** и убеждаемся, что задан метод вращения варимакс (**Method: Varimax**). Нажимаем **Continue**.

B) *Для вычисления факторных оценок* нажимаем **Scores...** (Оценки...) и отмечаем флажком **Save as variables** (Сохранить как переменные) и **Display factor coefficient matrix** (Выводить матрицу коэффициентов факторных оценок). Нажимаем **Continue**. Нажимаем OK.

11. Получаем результаты. К тем результатам, которые были описаны в п. 9, добавляются новые:

A) Матрица коэффициентов факторных оценок:

Factor Score Coefficient Matrix

	Factor		
	1	2	3
v1	.060	.113	-.493
v2	-.018	.158	-.045

	Factor		
	1	2	3
v3	.005	-.052	.408
v4	.095	.055	-.063
v5	.046	-.204	-.078
v6	-.020	.266	.128
v7	-.696	.172	-.051
v8	.180	-.121	-.002
v9	.006	.500	-.192
v10	.084	.085	-.012

Extraction Method: Maximum Likelihood. Rotation Method: Varimax with Kaiser Normalization. Factor Scores Method: Regression.

Эти коэффициенты являются стандартизированными регрессионными β-коэффициентами трех уравнений множественной регрессии — по одному для каждого фактора. Каждый из них пропорционален соответствующей факторной нагрузке и обратно пропорционален корреляции данной переменной с другими переменными. Факторные оценки по каждому фактору для объекта получаются путем подстановки в регрессионное уравнение с указанными β-коэффициентами z-значений соответствующих исходных переменных.

В) Новые переменные в составе таблицы исходных данных (SPSS Data Editor).

После выполнения пп. 10–В в таблице исходных данных **Data Editor** после исходных переменных появляются новые переменные, соответствующие вычисленным факторным оценкам для объектов. В данном случае для каждого испытуемого вычислены по три значения факторных оценок: FAC1_1, FAC2_1, FAC3_1. Эти переменные могут быть использованы для дальнейшего анализа вместо набора исходных переменных.

Глава 17

ДИСКРИМИНАНТНЫЙ АНАЛИЗ

НАЗНАЧЕНИЕ

Дискриминантный анализ представляет собой альтернативу множественного регрессионного анализа для случая, когда зависимая переменная представляет собой не количественную (номинативную) переменную. При этом дискриминантный анализ решает, по сути, те же задачи, что и множественный регрессионный анализ (МРА): предсказание значений «зависимой» переменной, в данном случае — категорий номинативного признака; определение того, какие «независимые» переменные лучше всего подходят для такого предсказания. *Структуры исходных данных* для дискриминантного и множественного регрессионного анализа практически идентичны:

№	X_1	X_2	...	X_P	Y
1	x_{11}	x_{12}	...	x_{1P}	y_1
2	x_{21}	x_{22}	...	x_{2P}	y_2
...	...	...	...	...	...
N	x_{N1}	x_{N2}	...	x_{NP}	y_N

Строки этой таблицы соответствуют объектам (испытуемым), а столбцы — переменным. Переменные x_1, ..., x_P представлены в количественной шкале. *Различие* исходных данных для дискриминантного и множественного регрессионного методов заключается лишь в том, что представляет собой «зависимая» переменная Y: для МРА она является *количественной*, а для дискриминантного анализа — *номинативной* (классифицирующей) переменной.

В то же время дискриминантный анализ можно определить и как метод *классификации*, так как «зависимая» переменная — номинативная, то есть она классифицирует испытуемых на группы, соответствующие разным ее градациям. В этом смысле *исходными данными* для дискриминантного анализа является группа N объектов (испытуемых), разделенная на G классов так, что каждый объект отнесен к одному и только одному классу (градации номинативной переменной). Допускается при этом, что некоторые объекты не отне-

сены к какому-либо из этих классов (являются «неизвестными»). Для каждого из объектов имеются данные по *P* количественным признакам, одним и тем же для этих объектов. Эти количественные признаки называются **дискриминантными переменными**. *Задачами дискриминантного анализа* являются: определение решающих правил, позволяющих по значениям дискриминантных переменных отнести каждый объект (в том числе и «неизвестный») к одному из известных классов; определение «веса» каждой дискриминантной переменной для разделения объектов на классы.

ПРИМЕР

В качестве объектов могут выступать студенты, сгруппированные по успешности обучения, а в качестве дискриминантных переменных — результаты их вступительных испытаний, социально-демографические характеристики и пр. При помощи дискриминантного анализа мы можем выделить переменные, наиболее важные для предсказания успешности обучения. Кроме того, по этим показателям мы можем предсказать успешность обучения абитуриентов.

Испытуемыми могут быть клиенты психотерапевта, сгруппированные по эффекту оказанной помощи. Переменными — симптомы, различные социальные и психологические показатели, а также характеристики видов помощи (длительность и характер терапии и пр.). При помощи дискриминантного анализа исследователь может определить переменные, наиболее существенные для эффекта психотерапии, а также предсказать результативность терапии для данного клиента при использовании данного вида помощи.

Таким образом, дискриминантный анализ позволяет решить две группы проблем:

1. Интерпретировать различия между классами, то есть ответить на вопросы: насколько хорошо можно отличить один класс от другого, используя данный набор переменных; какие из этих переменных наиболее существенны для различения классов. Сходную задачу решает дисперсионный анализ.

2. Классифицировать объекты, то есть отнести каждый объект к одному из классов, исходя только из значений дискриминантных переменных. Задача классификации связана с получением по данным об «известных» объектах дискриминантных функций «решающих правил», позволяющих по значениям дискриминантных переменных отнести с известной вероятностью каждый объект к одному из классов.

В решении задачи классификации дискриминантный анализ является не заменимым другими методами. Часто дискриминантный анализ называют еще «классификацией с обучением» или «распознаванием образов». В первом случае предполагают, что мы «учимся» классифицировать «неизвестные» объекты по дискриминантным переменным, используя данные об «известных» объектах. Во втором случае под «образом» объекта подразумевается совокупность измеренных для него значений дискриминантных переменных. И дискриминантный анализ позволяет в этом смысле распознать образ «нового» объекта путем отнесения его к известному классу объектов.

Дискриминантный анализ имеет общие черты с многомерным дисперсионным анализом (MANOVA). По сути, дискриминантные переменные можно рассматривать как многомерную зависимую переменную, а классифицирующую переменную — как фактор. Этот подход применяется для определения достоверности различения классов по совокупности всех переменных (по λ-Вилкса) и по каждой из дискриминантных переменных в отдельности (по критерию *F*-Фишера) — как в дисперсионном, так и в дискриминантном анализе.

Сравнивая дискриминантный и множественный регрессионный анализ, можно отметить их сходство в отношении решаемой задачи — предсказания. Однако дискриминантный анализ, являясь более сложным методом, имеет свои преимущества. В качестве «зависимой» переменной в дискриминантном анализе выступает классификация, что делает метод более универсальным: любое измерение можно свести к шкале наименований и избежать требования нормальности распределения «зависимой» переменной. Прогностическая эффективность дискриминантого анализа обычно выше, чем МРА, так как для предсказания используется не одна функция, как в МРА, а, как правило, несколько. Наконец, дискриминантный анализ позволяет провести более глубокое исследование различий между градациями «зависимой» переменной и влияния на нее «независимых» переменных.

МАТЕМАТИКО-СТАТИСТИЧЕСКИЕ ИДЕИ МЕТОДА

Классы, на которые разбито множество объектов, можно представить как значения некоторой *классифицирующей* («зависимой») *переменной*, измеренной в шкале наименований. *Дискриминантные переменные* представлены в числовой шкале. *Основная задача* дискриминантного анализа заключается в том, чтобы по значениям дискриминантных переменных для объектов получить значения классифицирующей переменной, то есть определить классы, в которые попадают эти объекты.

Дискриминантные переменные, количество которых равно *P*, можно представить себе как ортогональные оси *p*-мерного евклидова пространства. Тогда каждый объект будет являться точкой в этом пространстве, положение которой задано значениями дискриминантных переменных для этого объекта как его координатами. Так, если переменных две, то объект может быть изображен на плоскости в месте пересечения координат, соответствующих значениям этих двух переменных для данного объекта. Если переменных три, то объект представляет собой точку в трехмерном пространстве, и т. д.

Множество объектов в пространстве *P* признаков можно представить как скопление точек. Чем более объекты похожи друг на друга по данным признакам, тем плотнее будет скопление точек. Если несколько классов объектов отличаются друг от друга по дискриминантным переменным, то их мож-

но представить как соответствующие классам скопления точек в некоторых областях P-мерного пространства признаков. Чем больше объекты внутри каждого класса похожи друг на друга и отличаются от объектов из другого класса, тем меньше пересечений соответствующих классам «территорий».

Для каждого класса в пространстве признаков можно определить положение центроида — точки, координаты которой есть средние значения дискриминантных переменных для данного класса. **Центроид** — это место типичных наблюдений для данного класса, его можно использовать как для описания различий между классами, так и для определения принадлежности «неизвестных» объектов к одному из классов.

Из геометрической интерпретации задачи дискриминантного анализа следует *правило классификации объектов*: объект приписывается к тому классу, к центроиду которого он ближе всего. Соответственно, сама задача классификации объектов сводится к определению расстояний от каждого объекта до центроидов каждого класса по известным значениям дискриминантных переменных.

В современных компьютерных программах задача классификации решается с помощью канонических дискриминантных функций. **Канонические функции** — это ортогональные оси, в максимальной степени различающие центроиды классов. Началом координат для канонических функций является «главный центроид» — точка, координаты которой есть средние значения всех дискриминантных переменных. Первая каноническая ось ориентирована в направлении, в котором центроиды классов различаются в максимальной степени. Если классов больше двух, то вторая ось ориентирована перпендикулярно первой в направлении максимального различия классов и т. д. Максимальное число таких функций равно числу классов за вычетом единицы. Так, для различения двух центроидов (классов) достаточно одной оси, для различения трех классов — двух канонических функций, и т. д. Таким образом, канонические функции позволяют преобразовать P-мерное пространство исходных признаков в Q-мерное пространство дискриминантных функций ($Q = G - 1$, где G — число классов). Обсуждение процедуры получения канонических функций выходит за рамки этой книги. Отметим лишь, что в ее основе лежит анализ ковариационных и корреляционных матриц, а процедура их получения и результат весьма напоминают факторный анализ.

Канонические функции и дискриминантные переменные связывают **стандартизированные канонические коэффициенты**, которые позволяют оценить относительный вклад переменных в каждую каноническую функцию. В отличие от них, **структурные коэффициенты** канонических функций — это корреляции канонических функций и дискриминантных переменных. Как и факторные нагрузки в факторном анализе, *структурные коэффициенты* отражают связь дискриминантных переменных с каноническими функциями. Структурные коэффициенты канонических функций показывают *вклад* каждой дискриминантной переменной в различительную способность соответствующей функции. Таким образом, каждая каноническая функция может быть

проинтерпретирована через переменные, вносящие в нее наибольший по абсолютной величине вклад — подобно интерпретации факторов по факторным нагрузкам в факторном анализе.

Анализ канонических функций сопровождается получением важных статистических показателей качества классификации. Основными из них являются: *собственное значение канонической функции*, *λ-Вилкса* и *χ^2-тест*.

Собственное значение канонической функции, как и в факторном анализе, есть показатель информативности функции. Сумма всех собственных значений равна числу классов. Соответственно, собственное значение для данной канонической функции, деленное на количество классов, есть показатель ее информативности — доли суммарной дисперсии всех объектов по всем переменным, которая исчерпывается этой канонической функцией.

λ-Вилкса выполняет ту же функцию, что и в MANOVA, то есть является мерой достоверности различения классов при помощи данного набора переменных. **λ-Вилкса** — это мера остаточной дискриминативной способности переменных при учете данного набора канонических функций. Следовательно, чем меньше λ-Вилкса, тем лучше данная каноническая функция (или весь их набор) различает объекты. **χ^2-тест** позволяет определить статистическую достоверность такого различения.

ПРИМЕР

Предположим, в результате дискриминантного анализа для трех классов были получены две канонические функции. Основные их показатели приведены ниже в таблице.

Функции	Собственное значение	% дисперсии	λ-Вилкса	χ^2	*p*-уровень
1	2,794	95,6	0,233	22,549	0,004
2	0,129	4,4	0,886	1,879	0,598

Первая каноническая функция обладает 95,6% общих дискриминативных возможностей, а вторая — всего 4,4%. Величина λ = 0,233 для первой канонической функции показывает остаточную дискриминативную способность после учета всех канонических функций, а величина λ = 0,886 — остаточную дискриминативную способность при учете только второй канонической функции. Общая дискриминативная способность канонических функций достоверна на высоком уровне статистической значимости ($p = 0,004$), а статистическая значимость второй канонической функции явно мала ($p = 0,598$). Таким образом, различие классов по второй канонической функции не подлежит содержательной интерпретации. В принципе, ее можно исключить из анализа, но при условии, что качество классификации при этом сохранится на приемлемом для исследователя уровне.

Значения канонических функций вычисляются для каждого объекта по формуле, которая идентична по виду линейному уравнению множественной регрессии:

$$Y_{ik} = b_{k0} + b_{k1}x_{1i} + b_{k2}x_{2i} + \ldots + b_{kP}x_{Pi},$$

где Y_{ik} — *значение канонической функции* k для объекта i, а $b_{k0}, ..., b_{kP}$ — *канонические коэффициенты* для каждой из дискриминантных переменных. *Значения канонических функций* вычисляются для каждого центроида и каждого объекта, в том числе — «неизвестного», для которого не известна принадлежность к классу, и интерпретируются как их координаты в пространстве канонических функций. В этом пространстве малой размерности можно получить наглядное отображение всех объектов вместе с центроидами классов.

Принадлежность объекта к классу в большинстве компьютерных программ дискриминантного анализа определяется по *расстоянию* этого объекта до центроида соответствующего класса в пространстве канонических функций. Объект причисляется к тому классу, к центроиду которого он ближе всего. Однако надо помнить, что если расстояние объекта до класса велико (то есть профиль объекта мало похож на среднегрупповой), то объект может быть причислен к данному классу, поскольку до остальных классов он еще дальше.



Определение принадлежности неизвестных объектов

Производной от расстояния является еще одна мера классификации — *апостериорная вероятность* принадлежности к классу. *Априорная вероятность* («до опыта») принадлежности «нового» объекта к классу равна численности «известных» объектов этого класса, деленной на все «известные» объекты. Эта вероятность известна и без дискриминантного анализа, «до опыта». **Апостериорная вероятность** («после опыта») вычисляется исходя из расстояний данного объекта до центроидов каждого класса в предположении, что он принадлежит к одному из этих классов. Для любого объекта, следовательно, сумма этих вероятностей по всем классам равна 1. И чем меньше расстояние этого объекта до центроида класса, тем выше апостериорная вероятность его принадлежности к этому классу. Отнесение объекта к классу на основе наибольшей из вероятностей, таким образом, эквивалентно использованию наименьшего расстояния до центроида этого класса.

Вычисленные расстояния или апостериорные вероятности для известных объектов позволяют определить точность классификации и проанализировать ошибки, а для неизвестных — отнести объекты к одному из классов.

Анализ дискриминантных переменных позволяет, если это необходимо, отсеять несущественные для предсказания дискриминантные переменные. Наиболее важными показателями в этом анализе являются: *критерий F-Фишера*, *толерантность* и *статистика F-удаления*. Значимость каждой переменной для разделения классов определяется по F-Фишера по модели дисперсионного анализа. *Толерантность* равна единице минус квадрат коэффициента

множественной корреляции этой переменной со всеми остальными. Если толерантность равна нулю, то эта переменная является линейной комбинацией одной или нескольких других переменных и ее нельзя включать в анализ, равно как и переменные с очень малой толерантностью (скажем, меньше 0,001). **Статистика *F*-удаления** оценивает ухудшение разделения классов при удалении данной переменной из набора. Следовательно, чем больше значение этой статистики, тем более значима данная переменная для различения классов. На величину статистики *F*-удаления влияет не только различительная способность самой этой переменной (как в модели дисперсионного анализа), но и ее связь с другими переменными: чем сильнее она связана с другими переменными, тем меньше статистика *F*-удаления, тем меньше значение данной переменной.

Компьютерные программы позволяют автоматически отсеять малозначимые для дискриминантного анализа переменные. Во-первых, программа (SPSS) автоматически исключает из анализа переменные с низкой толерантностью. Во-вторых, возможен **пошаговый дискриминантный анализ**. При пошаговом методе переменные удаляются из анализа или включаются в него на основе улучшения (ухудшения) качества различения классов (обычно — по λ-Вилкса). Критериями для включения и удаления переменной являются статистики *F*-включения и *F*-удаления, которые показывают степень улучшения и ухудшения различения классов при включении и удалении данной переменной. Численные значения этих статистик могут быть заданы пользователем программы.

Дополнением к задаче классификации является *анализ расстояний между классами*. Программы обычно вычисляют значения *F*-критерия Фишера и *р*-уровень статистической значимости расстояния. Анализ расстояний позволяет определить, насколько существенно различаются классы по выбранным для анализа дискриминантным переменным.

Несмотря на обилие статистических критериев и показателей качества классификации, основным ориентиром для исследователя должно все же являться сопоставление действительной классификации «известных» объектов и их классификации при помощи канонических функций. Таким образом, *основным показателем качества* является процент совпадения этих двух классификаций.

Дискриминантный анализ относится к наиболее сложным методам, поэтому здесь мы ограничились лишь минимумом сведений, необходимых для понимания его основ. Более детальную информацию о порядке интерпретации основных показателей дискриминантного анализа можно почерпнуть в последнем разделе этой главы, где разбирается конкретный пример применения этого метода[1].

---

[1] Подробное изложение математико-статистических идей метода можно найти в книге: Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. М., 1989.

ИСХОДНЫЕ ДАННЫЕ И ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Исходными для дискриминантного анализа являются *N* объектов (испытуемых), разбитых на *G* классов так, чтобы в каждом классе содержалось не менее двух объектов. Для каждого из них имеются данные по *P* переменным. В отношении количества переменных строгих ограничений нет, но часто рекомендуют не менее чем двукратное превышение числа объектов над числом переменных.

Исходные данные удобно представлять в виде матрицы размерностью *N* строк и *P*+1 столбцов, где добавленная переменная содержит обозначение принадлежности объекта к классу. Если у нас, предположим, имеется 3 класса, то каждый «известный» объект маркируется в этом столбце цифрой 1, 2 или 3 — в зависимости от принадлежности к классу. «Неизвестные» объекты при этом не обозначаются.

Статистические принципы, заложенные в основу анализа, предполагают измерение дискриминантных переменных в метрической шкале (интервалов или отношений). Кроме того, предполагается нормальное распределение каждого признака. Если эти требования не выполняются, использование статистических показателей классификации становится некорректным. Тем не менее, исследователь может отступать от этих предположений, определяя предсказательную силу дискриминантного анализа по соотнесению действительной и предсказанной классификации для «известных» объектов.

Другое требование к исходным данным — отсутствие линейных зависимостей между дискриминантными переменными (коэффициентов корреляции, равных 1). Мера соблюдения этого требования — толерантность каждой переменной — вычисляется в ходе анализа. Если толерантность какой-либо переменной слишком низка, компьютерная программа (в SPSS) автоматически ее исключает.

ПРИМЕР 17.1

Предположим, что, как и в примере для МРА, перед исследователем стоит задача предсказания успеваемости пяти абитуриентов по данным вступительных тестов (4 теста). Кроме того, его интересует, какие тесты, как дискриминантные переменные, обладают наибольшей предсказательной силой в отношении последующей успеваемости. В качестве исходных данных психолог имеет для каждого из 20 учащихся предыдущего набора 4 показателя тестирования. Кроме того, ему известно, к какой группе успешности принадлежит каждый из этих студентов: неуспевающих (0), средних (1) и отличников (2). В его



распоряжении имеются результаты применения тех же 4 тестов для пяти абитуриентов, и исследователь надеется предсказать принадлежность каждого из них к одной из трех групп. Таким образом, исходными данными для дискриминантного анализа являются: классифицирующая переменная — принадлежность к одной из трех групп (Y) и 4 дискриминантных переменных — результатов тестов (test 1, test 2, test 3, test 4). Исходные данные представляют собой следующую таблицу:

Таблица 17.1

Пример исходных данных для дискриминантного анализа

№	test 1	test 2	test 3	test 4	Y
1	86,00	110,00	110,00	101,00	1
2	62,00	97,00	99,00	100,00	0
3	93,00	107,00	103,00	103,00	1
4	87,00	117,00	93,00	95,00	0
...	...	...	...	...	...
20	150,00	118,00	107,00	110,00	2
21	96,00	114,00	114,00	103,00	
22	104,00	73,00	105,00	95,00	
23	94,00	121,00	115,00	104,00	
24	91,00	129,00	105,00	98,00	
25	74,00	121,00	100,00	100,00	

(Для абитуриентов 21–25 значения классифицирующей переменной Y не определены).

Основные результаты дискриминантного анализа:

1. Определение статистической значимости различения классов при помощи данного набора дискриминантных переменных. Показатели — λ-Вилкса, χ^2-тест, p-уровень значимости.
2. Классификация «известных» и «неизвестных» объектов при помощи расстояний или значений априорных вероятностей. Качество классификации определяется совпадением действительной классификации и предсказанной для «известных» объектов. Мерой качества может служить вероятность ошибочной классификации как соотношение количества ошибочного отнесения к общему количеству «известных» объектов.
3. Выяснение вклада каждой переменной в дискриминантный анализ. Определяется по значениям критерия F-Фишера, толерантности и статистики F-удаления.
4. Вычисление расстояний между центроидами классов и определение их статистической значимости по F-критерию.
5. Анализ канонических функций, их интерпретация через дискриминантные переменные (по стандартизированным и структурным коэффициентам канонических функций).
6. Графическое представление всех объектов и центроидов классов в осях канонических функций.

Рассмотрим процедуру обработки, основные результаты и их интерпретацию, применив программу SPSS к данным примера 17.1.

ОБРАБОТКА НА КОМПЬЮТЕРЕ

Исходные данные (**Data Editor**) представляют собой 5 переменных (`test1`, `test2`, `test3`, `test4`, `Y`) для 25 объектов; для последних 5 объектов значения *Y* не определены (табл. 17.1).

1. Выбираем **Analyze > Classify** (Классификация) **> Discriminant**... (Дискриминантный).

2. В открывшемся основном окне диалога **Discriminant Analysis** (Дискриминантный анализ) выделяем и переносим из левого окна переменные при помощи кнопки ▷: независимые переменные (`test1`, `test2`, `test3`, `test4`) — в правое второе сверху окно (**Independents**), классифицирующую переменную (`Y`) в правое верхнее окно (**Grouping Variable**). Нажимаем кнопку **Define range...** (Определить классы...) и задаем минимальное значение (**Minimum:** `0`) и максимальное значение (**Maximum:** `2`) номеров классов. Нажимаем **Continue**.

3. В том же окне диалога выбираем метод. Для этого вместо принятого по умолчанию метода **Enter independents together** (Введение всех независимых переменных) при помощи правой кнопки мыши отмечаем **Use stepwise method** (Применить пошаговый метод). При этом высвечивается кнопка **Method...** (Метод). Нажимаем эту кнопку. В открывшемся окне диалога пошагового метода отмечаем флажком *F*-критерий для расстояний между классами (**F for pairwise distances**). В поле **Method** (Метод) оставляем принятый по умолчанию метод λ-Вилкса (**Wilk's lambda**). Меняем принятые по умолчанию *F*-критерии включения и удаления: для включения устанавливаем 0,1 (**Entry:** `0,1`), для удаления устанавливаем 0 (**Removal:** `0`). Нажимаем **Continue**.

Примечание. В данном случае мы выбираем пошаговый метод не столько для того, чтобы отсеять несущественные переменные, сколько для получения более подробных результатов, так как при установках по умолчанию программа SPSS не позволяет получить ряд важных показателей. Минимально возможные значения *F*-критериев для включения и удаления переменных устанавливаем, чтобы получить пошаговые результаты вплоть до полного набора переменных.

4. В том же окне нажимаем кнопку **Statistics...** (Статистики...) и отмечаем **Univariate ANOVAs** (Однофакторные ANOVA) — для получения достоверности различия классов по каждой переменной в отдельности, по критерию *F*-Фишера. Нажимаем **Continue**.

5. Для уточнения результатов классификации нажимаем кнопку **Classify...** (Классификация...) и отмечаем **Prior probabilities: Compute of group size** (Априорные вероятности: исходя из размера группы) вместо принятого по умолча-

нию **All groups equal** (Все группы равны по численности). В поле **Display** (Показывать) отмечаем флажком **Casewise results** (Результаты для всех объектов) — для получения результатов классификации для всех объектов. В поле **Plots** (Графики) отмечаем флажком **Combined groups** (Совмещенные группы) — для графического изображения всех объектов в осях канонических функций. Нажимаем **Continue**.

Примечание. В данном случае мы предполагаем, что соотношение численности групп (отличников, успевающих и неуспевающих студентов) заранее известно исходя из имеющегося соотношения. Если это не так, то разумнее принять версию о равной численности групп.

6. Для вычисления и сохранения оценок классифицирующей переменной и апостериорных вероятностей в том же окне диалога нажимаем клавишу **Save...** (Сохранить). В появившемся окне диалога отмечаем флажком **Predicted group membership** (Предсказанная принадлежность к группе) и **Probabilities of group membership** (Вероятность принадлежности к группам). Нажимаем **Continue** (Продолжить).

Примечание. Этот этап — сохранение результатов классификации — имеет смысл выполнять на последнем этапе анализа, когда выбрана окончательная модель предсказания.

7. После указания всех установок в основном окне диалога **Discriminant Analysis** (Дискриминантный анализ) нажимаем ОК и получаем результаты.

Рассмотрим основные результаты дискриминантного анализа.

Поскольку обычно исследователя интересует общая эффективность предсказания, то лучше обратиться к соответствующим числовым результатам (статистикам) для каждого объекта:

А) Статистики для объектов:

Casewise Statistics

Case Number	Actual Group	Predicted Group
1	1	1
2	0	0
3	1	1
4	0	0
5	1	0(**)
6	0	0
7	0	0
8	2	1(**)
9	2	2
10	1	1
11	0	0
12	2	2

Case Number	Actual Group	Predicted Group
13	1	1
14	0	0
15	2	2
16	2	2
17	1	1
18	2	2
19	1	1
20	1	1
21	ungrouped	1
22	ungrouped	1
23	ungrouped	0
24	ungrouped	2
25	ungrouped	2

** Misclassified case

Эта таблица содержит указания на принадлежность каждого объекта к группе: действительную (Actual) и предсказанную на последнем шаге анализа (Predicted). Предсказанные значения приведены и для последних пяти «неизвестных» объектов. Ошибочные предсказания для двух объектов отмечены «звездочкой». Таким образом, точность предсказания составляет 90%, то есть очень высока.

Те же данные плюс значения вероятностей принадлежности каждого объекта к каждому классу сохранены как новые переменные *в таблице исходных данных* (Data Editor):

№	test 1	test 2	test 3	test 4	*Y*	Dis_1	Dis1_1	Dis1_2	Dis1_3
1	86,00	110,00	110,00	101,00	1	1	,044	,840	,116
...	...	...	...	...	...	...	...	...	...
24	91,00	129,00	105,00	98,00	,	2	,001	,376	,623
25	74,00	121,00	100,00	100,00	,	2	,012	,492	,496

В исходных данных появляются новые переменные: Dis_1 — предсказанная принадлежность каждого объекта к одному из классов, Dis1_1, Dis1_2, Dis1_3 — значения апостериорных вероятностей принадлежности каждого объекта к каждому классу. Значения вероятностей позволяют судить, насколько уверенно объекты могут быть отнесены к тому или иному классу. Так, объект № 25 почти равновероятно может быть отнесен и ко второму (1) и к третьему (2) классу, а объект № 1 уверенно отнесен ко второму (1) классу.

В) Результаты ANOVA для каждой дискриминантной переменной (см. табл. Tests of Equality of Group Means).

Эти результаты позволяют отобрать для дискриминантного анализа только те переменные, различия по которым между группами статистически до-

Tests of Equality of Group Means

	Wilks' Lambda	F	df1	df2	Sig.
test1	.629	5.017	2	17	.019
test2	.669	4.212	2	17	.033
test3	.683	3.952	2	17	.039
test4	.569	6.444	2	17	.008

стоверны. В данном случае обнаружены статистически значимые различия групп по каждой из дискриминантных переменных. Более детальный результат можно получить, проведя многомерный ANOVA в отношении дискриминантных переменных (см. главу 13). Помимо достоверности различия всех групп по каждой переменной, полезны будут множественные сравнения — для изучения попарных различий между группами по этим переменным.

С) Показатели пошагового анализа.

(Stepwise Statistics) — это результаты, которые можно получить только при помощи пошагового метода.

Дискриминантные переменные для каждого шага

Variables in the Analysis

Step		Tolerance	F to Remove	Wilks' Lambda
1	test4	1.000	6.444	
2	test4	.864	7.922	.669
	test2	.864	5.543	.569
3	test4	.863	5.223	.439
	test2	.850	4.873	.427
	test1	.982	2.231	.336
4	test4	.665	2.677	.323
	test2	.850	4.423	.381
	test1	.974	2.150	.305
	test3	.743	.764	.259

В таблице указаны толерантность, значения статистики *F*-удаления и λ-Вилкса для каждой переменной на каждом шаге анализа. Видно, что переменная test 3 наименее значима для анализа, и ее, в случае необходимости, можно исключить.

Значения λ-Вилкса для каждого шага анализа

Step	Number of Variables	Lambda	df1	df2	df3	Exact F			
						Statistic	df1	df2	Sig.
1	1	.569	1	2	17	6.444	2	17.000	.008
2	2	.336	2	2	17	5.802	4	32.000	.001
3	3	.259	3	2	17	4.826	6	30.000	.001
4	4	.233	4	2	17	3.744	8	28.000	.004

λ-Вилкса показывает, что на каждом шаге увеличивается дискриминативная способность набора дискриминантных переменных (λ уменьшается). Однако уменьшение λ-Вилкса от 3-го к 4-му шагу минимально, а статистическая значимость различения даже падает (с $p = 0{,}001$ до $p = 0{,}004$). Следовательно, присоединение переменной test 3 (на последнем шаге), несущественно увеличивая дискриминативную способность набора переменных, приводит к уменьшению статистической значимости решения. Это еще больше убеждает в возможности удаления переменной test 3 из анализа. Однако для окончательного принятия решения об удалении переменных необходимо проверить, насколько при этом уменьшается точность классификации (по предсказанной классификации).

Попарное сравнение групп.

Pairwise Group Comparisons (*a*, *b*, *c*, *d*)

Step	y		0	1	2
1	0	F		2.320	12.686
		Sig.		.146	.002
	1	F	2.320		5.219
		Sig.	.146		.035
	2	F	12.686	5.219	
		Sig.	.002	.035	
2	0	F		2.689	15.503
		Sig.		.098	.000
	1	F	2.689		6.606
		Sig.	.098		.008
	2	F	15.503	6.606	
		Sig.	.000	.008	
3	0	F		3.306	13.440
		Sig.		.049	.000
	1	F	3.306		4.761
		Sig.	.049		.016
	2	F	13.440	4.761	
		Sig.	.000	.016	
4	0	F		2.893	9.774
		Sig.		.061	.001
	1	F	2.893		3.346
		Sig.	.061		.040
	2	F	9.774	3.346	
		Sig.	.001	.040	

a 1, 17 degrees of freedom for step 1.
b 2, 16 degrees of freedom for step 2.
c 3, 15 degrees of freedom for step 3.
d 4, 14 degrees of freedom for step 4.

В этой таблице для каждого шага приведены результаты проверки статистической достоверности различия групп друг от друга. Видно, что статистическая значимость различия всех трех групп друг от друга достигается на шаге 3 (без переменной test 3), а на шаге 4 статистическая достоверность, при включении этой переменной, даже падает.

По результатам пошагового анализа возникает впечатление, что переменная test 3 ничего не добавляет в различение классов. Однако попытка ее удаления из анализа приводит к тому, что количество ошибок при классификации увеличивается в два раза и составляет 20%. Поэтому целесообразно принять решение о сохранении всех дискриминантных переменных.

Следует отметить, что такое решение справедливо в отношении именно этих данных. При наличии большого количества дискриминантных переменных следует смелее удалять малозначимые переменные, добиваясь разумного компромисса между статистической значимостью результата и точностью предсказания.

D) Итоги анализа канонических дискриминантных функций (Summary of Canonical Discriminant Functions).

Собственные значения для канонических функций

Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	2.794(a)	95.6	95.6	.858
2	.129(a)	4.4	100.0	.338

a First 2 canonical discriminant functions were used in the analysis.

Первая каноническая функция более чем в 20 раз более информативна (95,6%), чем вторая функция (4,4%).

Лямбда Вилкса для каждой функции

Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1 through 2	.233	22.549	8	.004
2	.886	1.879	3	.598

Первая строка содержит значение $\lambda = 0{,}233$ и статистическую значимость $p = 0{,}004$ для всего набора канонических функций, вторая строка содержит данные для оставшейся дискриминантной способности набора после исключения первой функции. В данном случае полный набор обладает очень высокой дискриминантной способностью, которая резко падает после исключения первой канонической функции. Отсутствие статистической значимости второй дискриминантной функции означает, что ее интерпретация в отношении генеральной совокупности сомнительна.

Значения канонических функций для групповых центроидов.

Functions at Group Centroids

y	Function	
	1	2
0	-1.955	-.282
1	-.051	.405
2	2.023	-.259

Unstandardized canonical discriminant functions evaluated at group means.

В этой таблице приведены координаты центроидов для всех групп. Они позволяют интерпретировать канонические функции относительно их роли в различении классов. На отрицательном полюсе расположен центроид для первой группы (0), на положительном — центроид для третьей (2) группы, а средняя (1) группа расположена между ними. То есть чем больше значение этой функции, тем выше вероятность высокой успешности обучения. На положительном полюсе второй функции расположена средняя (1) группа, на отрицательном — две остальные группы. То есть чем выше значение данной функции, тем больше вероятность средней успешности обучения.

Стандартизированные коэффициенты канонических функций

Standardized Canonical Discriminant Function Coefficients

	Function	
	1	2
test1	.538	.496
test2	.775	-.345
test3	.219	.922
test4	.673	-.849

Эти коэффициенты позволяют определить соотношение вкладов переменных в каждую из канонических функций. Так, в первую каноническую функцию примерно с равным вкладом входят три из четырех переменных (кроме test 3): чем больше значения этих переменных, тем больше значение функции (то есть тем выше вероятность успешного обучения). Для второй канонической функции наиболее существенны переменные test 3 и test 4. Чем выше значение переменной test 3 и ниже значение по переменной test 4, тем больше значение второй канонической функции, то есть тем выше вероятность средней успешности обучения.

Структурные коэффициенты канонических функций

Structure Matrix

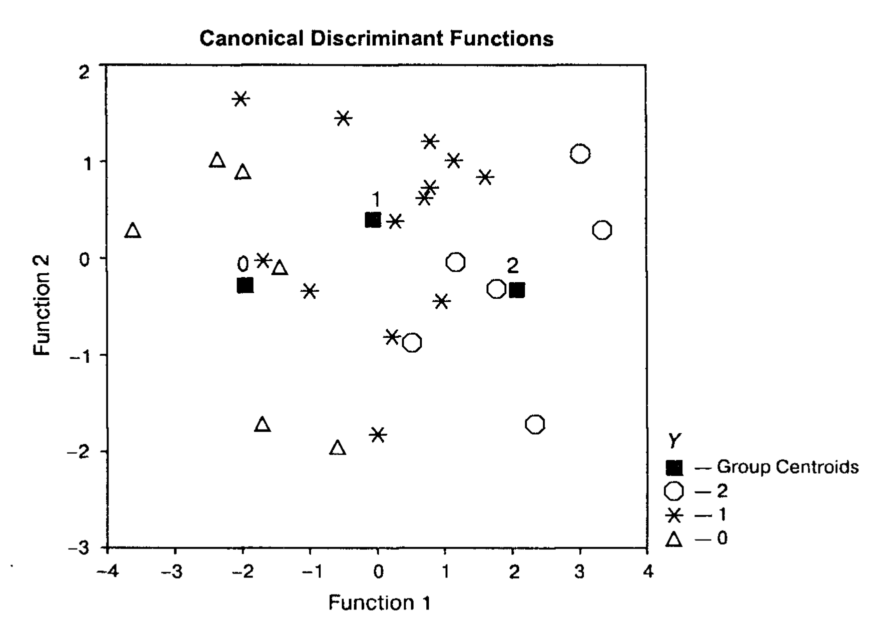
	Function	
	1	2
test4	.518(*)	-.240
test2	.418(*)	-.254
test3	.392	.526(*)
test1	.449	.451(*)

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions Variables ordered by absolute size of correlation within function.

* Largest absolute correlation between each variable and any discriminant function.

Как отмечалось, структурные коэффициенты, подобно факторным нагрузкам в факторном анализе, являются коэффициентами корреляции переменных с функциями. Следовательно, эти коэффициенты, как и в факторном анализе, позволяют интерпретировать канонические функции. Так, первая функция наиболее тесно связана с переменными test 4 и test 2: чем больше значения этих переменных, тем больше значение функции (то есть тем выше вероятность успешного обучения). Вторая функция наиболее тесно связана с переменными test 3 и test 1: чем больше значения этих переменных, тем больше вероятность средней успешности обучения, а чем меньше значения этих переменных, тем более вероятно, что студент будет либо отличником, либо неуспевающим.

Наконец, в соответствии с отмеченным флажком графиком для совмещенных групп, получаем *графическое изображение канонических дискриминантных функций.*



На графике изображены групповые центроиды (Group Centroids) и объекты в осях канонических функций. График помогает интерпретировать канонические функции и визуально оценить качество классификации по плотности объектов внутри каждого класса и по отчетливости границ между классами.

Глава 18

МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ

НАЗНАЧЕНИЕ

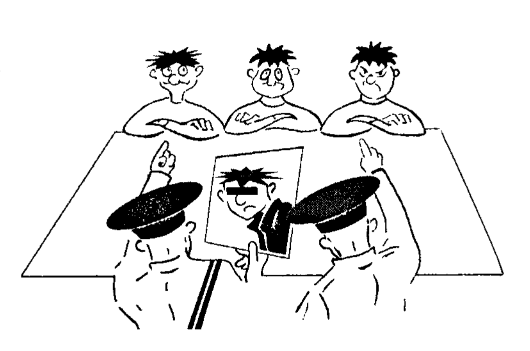
Основная *цель многомерного шкалирования* (МШ) — выявление структуры исследуемого множества объектов — близка к цели факторного и кластерного анализа. Так же, как в факторном анализе, под структурой понимается набор основных факторов (в данном случае — шкал), по которым различаются и могут быть описаны эти объекты. Однако в отличие от факторного, но подобно кластерному анализу исходной информацией для МШ являются данные о различии или близости объектов.

В психологии чаще всего исходными данными для МШ являются субъективные суждения испытуемых о различии или сходстве стимулов (объектов). Центральное положение МШ заключается в том, что в основе таких суждений лежит ограниченное число субъективных признаков (критериев), определяющих различение стимулов, и человек, вынося свои суждения, явно или неявно учитывает эти критерии. Основываясь на этом положении, решается *главная задача МШ* — реконструкция психологического пространства, заданного небольшим числом измерений-шкал, и расположение в нем точек-стимулов таким образом, чтобы расстояния между ними наилучшим образом соответствовали исходным субъективным различиям. Таким образом, **шкала** в МШ интерпретируется как критерий, лежащий в основе различий стимулов.

Геометрические представления МШ основаны на аналогии между понятием *различия* в психологии и понятием *расстояния* в пространстве. Чем более субъективно сходны между собой два объекта, тем ближе в реконструируемом пространстве признаков должны находиться соответствующие этим объектам точки. Исходя из такой *дистанционной модели*, по субъективным данным о различии одного объекта от другого реконструируется их взаимное расположение в пространстве нескольких признаков. Эти признаки трактуются как субъективные шкалы — критерии, которыми пользуется человек при различении объектов. А расстояние между объектами в этом пространстве есть определенная функция от исходных оценок различия.

Общая схема МШ формально может быть представлена следующим образом. На основе суждений экспертов (испытуемых) в отношении интересую-

Эксперты могут не только попарно сравнивать, но и упорядочивать объекты...



щих исследователя объектов вначале составляется симметричная **матрица попарных различий** (или матрицы — по одной для каждого эксперта). Допускается и использование **данных о предпочтениях**, содержащих упорядочивание каждым экспертом совокупности объектов по степени их предпочтения. Сравниваемыми объектами могут быть члены коллектива, предметы домашнего обихода, литературные отрывки, цветовые оттенки и т. д. *Модель* МШ предполагает, что эксперт производит сравнение, осознанно или нет пользуясь одним или несколькими признаками этих объектов. В отношении сотрудников подразделения такими признаками могут быть должностной статус, профессионализм, доброжелательность и т. д.

В процессе МШ определяется, сколько признаков-*шкал* необходимо и достаточно для построения координатного пространства и размещения в нем точек-объектов. Если δ_{ij} — это *оценка* экспертом *различия* между объектами i и j, а число *признаков*, которыми пользуется эксперт при сравнении, — K, то задача многомерного шкалирования сводится к определению всех x_{ik} и x_{jk} как координат этих объектов в пространстве K признаков. При этом предполагается, что число критериев, которыми пользуется эксперт, значительно меньше числа сравниваемых объектов. Если, например, i и j — сотрудники, а признак k — доброжелательность, то x_{ik} и x_{jk} — доброжелательность этих сотрудников. Важно отметить, что исследователю эмпирически даны только **оценки различий** δ_{ij}. Величины значений признаков x_{ik} и x_{jk} непосредственно не даны, но оцениваются в результате МШ в виде матрицы:

$$X = \begin{vmatrix} x_{11} & \dots & x_{1K} \\ \dots & \dots & \dots \\ x_{P1} & \dots & x_{PK} \end{vmatrix}$$

где P — количество сравниваемых объектов, K — количество шкал.

Элементы x_{ij} указанной матрицы рассматриваются как координаты P объектов в пространстве K признаков. Пространство определено так, что чем больше исходное различие между объектами, тем дальше друг от друга расположены объекты в этом пространстве. Каждая шкала результирующего пространства получает интерпретацию через объекты, находящиеся на противоположных полюсах шкалы.

Следует отметить, что исходными данными для МШ могут являться не только субъективные оценки различий, но и обычные данные типа «объект-признак». Но поскольку МШ предназначено для анализа различий, то для данных типа «объект-признак» необходимо, во-первых, определить, что бу-

дет подлежать шкалированию — сами объекты (строки) или признаки (столбцы). Во-вторых, необходимо задать *метрику различий* — то, как будут определяться различия между всеми парами изучаемых элементов. Проблема выбора мер различия обсуждается в следующем разделе данной главы[1].

Выбирая МШ, исследователь должен отдавать себе отчет в том, что это довольно сложный метод, применение которого к тому же связано с неизбежными потерями исходной информации о различии объектов. Поэтому, если задача исследования ограничивается классификацией объектов и нет оснований полагать, что эта классификация обусловлена *небольшим* числом независимых причин — критериев различий, то целесообразнее воспользоваться более простым методом — кластерным анализом (см. главу 19).

Рассмотрим исходные данные и основные результаты применения МШ на простом примере. Попытаемся, исходя из субъективных оценок расстояний между совокупностью объектов, реконструировать конфигурацию их взаимного расположения. Допустим, субъекту предъявляется 10 объектов, расположенных на плоскости в некоторой произвольной конфигурации, и дана инструкция оценить расстояние между каждым объектом и всеми остальными, присвоив 1 наименьшему расстоянию, 2 — следующему по величине и т. д. Примерно одинаковым расстояниям разрешим присваивать одинаковые числовые значения. В результате выполнения такого задания наблюдатель заполнил нижний треугольник матрицы попарных различий между объектами, в данном случае — расстояний (табл. 18.1). Можно ли восстановить исходную конфигурацию объектов по такой матрице различий? Оказывается, МШ справляется с подобными задачами. Применение программы неметрического МШ (программа SPSS) дает 2-шкальное решение (табл. 18.2).

Таблица 18.1

Субъективные оценки расстояний между 10 объектами

№	1	2	3	4	5	6	7	8	9	10
1	0									
2	1	0								
3	2	1	0							
4	3	2	1	0						
5	3	2	1	1	0					
6	3	2	2	2	1	0				
7	3	3	3	3	2	1	0			
8	2	2	2	3	2	1	1	0		
9	1	1	2	3	2	2	2	1	0	
10	2	1	1	2	1	1	2	1	1	0

[1] В связи с тем, что типичными исходными данными для МШ в психологии являются все же непосредственные оценки различий, изложение этой главы сопровождается примерами анализа исходной информации именно этого типа. Тем не менее подчеркнем, что для МШ допустимо применение и любых других исходных данных.

Таблица 18.2

Результаты МШ субъективных оценок расстояний между 10 объектами (по данным табл. 18.1)

№ объектов	Шкала 1 (Dim. 1)	Шкала 2 (Dim. 2)
1	0,932	−0,006
2	0,471	−0,302
3	0,019	−0,559
4	−0,471	−0,804
5	−0,497	−0,257
6	−0,493	0,263
7	−0,461	0,810
8	0,026	0,558
9	0,474	0,296
10	0,000	0,000

Каждая строчка таблицы — это координаты соответствующего объекта на плоскости. Графическое изображение всех 10 точек, в соответствии с табл. 18.2, приведено на рис. 18.1.

Взаимное расположение объектов в точности соответствует исходной конфигурации, предлагаемой наблюдателю (рис. 18.2). При этом обращает на себя внимание тот факт, что информация, полученная от наблюдателя, носит неметрический характер, так как расстояния оценивались по шкале порядка. Итоговая же конфигурация воспроизводит метрические соотношения в расположении объектов. Это связано с тем, что информация о различиях, содержащаяся в матрице субъективных оценок, хотя и является по сути порядко-

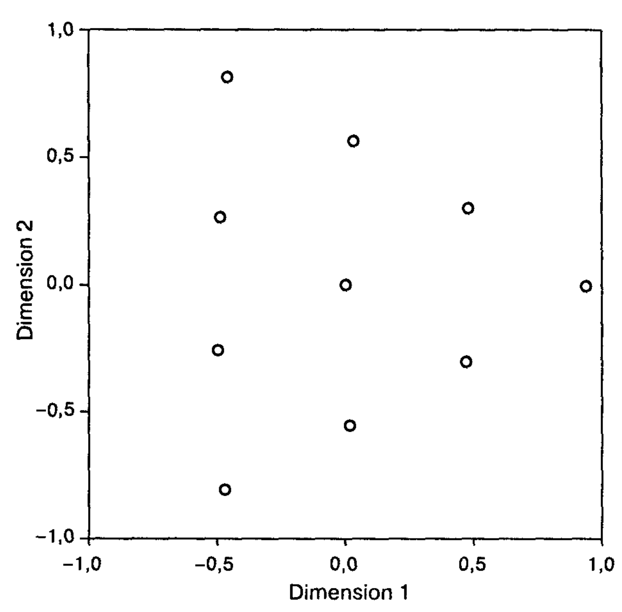


Рис. 18.1. Субъективное пространство 10 объектов по табл. 18.2

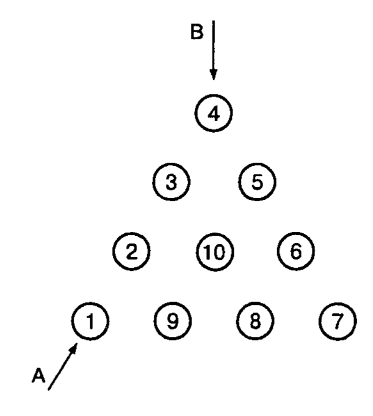


Рис. 18.2. Стимулы-объекты для наблюдателей А и В

вой, но обладает избыточностью, которая и позволяет восстановить метрические соотношения.

Сходные эксперименты проводились и с географическими данными (Терехина А. Ю., 1986). Например, испытуемым предлагалось оценить расстояния между 30 городами из различных частей земного шара, принимая за стандарт (100%) расстояние между Северным и Южным полюсами. Получаемые в результате применения МШ двумерные конфигурации в достаточной мере соответствуют истинному расположению выбранных городов на географической карте.

Если от каждого эксперта получена матрица попарных различий P объектов, то для таких данных используется *многомерное шкалирование индивидуальных различий*. При этом предполагается, что существует общее групповое пространство координат объектов в пространстве K общих признаков. Эксперты же отличаются тем, какой *«индивидуальный вес»* каждый из них придает тому или иному из K признаков при сравнении объектов. Соответственно, помимо групповой матрицы координат объектов, результатом этого варианта МШ является матрица индивидуальных весов размерностью $K \times N$.

Предположим, что объекты 1–10 (рис. 18.2) — объемные фигуры, расположенные на плоской поверхности в виде столбиков. Мы привлекаем двух наблюдателей А и В и располагаем их так, как показано на рисунке: чтобы они видели группу объектов на некотором отдалении вдоль плоскости, на которой расположены объекты, и не выше высоты объектов. Часть объектов при этом будет заслонять друг друга. Для наблюдателя А при этом не будут различимы пары объектов 1 и 10, 2–5, 9–6. Соответственно, для другого наблюдателя неразличимы пары 4–10, 3–9, 5–8.

Наблюдателям дается та же инструкция, что и в предыдущем примере: оценить расстояние между каждым объектом и всеми остальными, присваивая 1 наименьшему расстоянию, 2 — следующему за ним по величине и т. д. Расстоянию между невидимым объектом и объектом, их заслоняющим, наблюдатели присваивают значение 0 как неразличимым объектам. Результатом проведения такого испытания являются две матрицы попарных различий

Таблица 18.3

Результаты МШ индивидуальных различий восприятия 10 стимулов двумя наблюдателями

Stimulus Number	DIM.1	DIM.2
v1	–1,73	0,00
v2	–1,15	–0,58
v3	–0,58	–1,15
v4	0,00	–1,73
v5	0,58	–0,58
v6	1,15	0,58
v7	1,73	1,73
v8	0,58	1,15
v9	–0,58	0,58
v10	0,00	0,00
Subject Number	Subject Weights	
1 (*A*)	0,00	1,00
2 (*B*)	1,00	0,00

между 10 объектами, в каждой из которых по 3 ложных нулевых значения. К таким данным применима модель МШ индивидуальных различий. Результаты обработки (программа SPSS) приведены в табл. 18.3. 10 верхних строк этой таблицы — координаты объектов в двух шкалах, а две нижние строки — индивидуальные веса шкал.

Величина индивидуального веса по соответствующей шкале показывает, насколько учитывается признак при различении объектов. Как и следовало ожидать, каждый наблюдатель учитывает только одну координату — точку зрения на объекты, игнорируя другую. Тем не менее, двух точек зрения

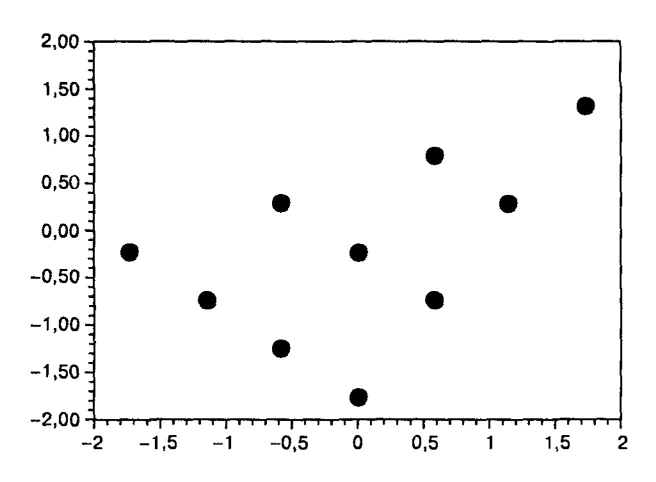


Рис. 18.3. Субъективное пространство 10 объектов по данным табл. 18.3

в данном случае достаточно, чтобы восстановить исходную конфигурацию объектов с точностью до порядковых отношений между расстояниями (рис. 18.3).

Метрические отношения между объектами не удалось воспроизвести точно в связи с малым числом наблюдателей. Если бы был третий наблюдатель, точка зрения которого отличалась от первых двух, то восстановились бы и метрические отношения.

Если эксперты не оценивают различия, а *упорядочивают объекты* по степени предпочтения, то применяется *модель предпочтения*. Исходными данными для МШ предпочтений является матрица размером $N \times P$, где N строк соответствуют экспертам, а P — это столбцы, содержащие ранговые места предпочтений объектов экспертами.

В дистанционной модели предпочтений каждый эксперт характеризуется координатами своей идеальной точки в пространстве K признаков. Эти координаты — такая комбинация характеристик объектов, которую эксперт считает идеальной. Результирующая матрица содержит, помимо координат объектов, и координаты идеальных точек — по одной для каждого эксперта.

Для нашего примера эксперты могли бы упорядочивать объекты по степени их наблюдаемой близости. Таким образом, описание объектов, полученное от каждого наблюдателя, стало бы экономнее. Результатом применения модели предпочтения была бы конфигурация объектов и «идеальных точек» наблюдателей, расположенных вблизи тех объектов, которые оказались рядом с наблюдателями.

Итак, МШ в своих основных трех модификациях позволяет решать три группы задач:

1. Исходные данные — прямые оценки субъектом различий между стимулами или вычисленные расстояния между объектами, характеризующимися совокупностью признаков. Примером второго типа данных могут являться расстояния между ролями (объектами), вычисленные по совокупности конструктов (репертуарные решетки Келли). МШ позволяет реконструировать психологическое пространство субъекта, как конфигурацию стимулов в осях существенных признаков, по которым эти стимулы различаются субъектом.

2. Исходные данные — те же, что и в предыдущем случае субъективные различия между стимулами (оцененные прямо или вычисленные), но полученные не от одного, а от группы субъектов. Взвешенная модель индивидуальных различий позволяет получить групповое психологическое пространство стимулов в осях общих для данной группы существенных признаков. Дополнительно к этому для каждого субъекта — индивидуальные веса признаков как меру учета соответствующих точек зрения при различении стимулов.

3. Исходные данные — результаты упорядочивания каждым из группы субъектов набора стимулов по степени предпочтения. Модель анализа предпочтений позволяет получить групповое психологическое пространство стимулов в осях существенных признаков и размещенные в этом же пространстве идеальные точки для каждого субъекта.

МЕРЫ РАЗЛИЧИЯ

Непосредственными данными для многомерного шкалирования является информация о различиях между объектами. Эти различия должны быть определены *между всеми парами объектов и иметь числовое выражение.* Первое требование означает, что анализируется матрица попарных различий $N \times N$ (где N — количество изучаемых объектов). Второе условие требует измерения различия, то есть введения *метрики.* Существуют четыре стандартных критерия, которым должна удовлетворять мера различия, чтобы быть метрикой:

1. *Симметрия.* Даны два объекта: x и y. Расстояние от x до y должно быть равно расстоянию от y до x.
2. *Неразличимость идентичных объектов.* Расстояние между двумя идентичными объектами равно 0.
3. *Различимость нетождественных объектов.* Расстояние между двумя различающимися объектами не равно 0.
4. *Неравенство треугольника.* Даны три объекта: x, y и z. Расстояние от x до y меньше или равно сумме расстояний от x до z и от z до y (длина любой стороны треугольника меньше или равна сумме двух других сторон).

Эти очевидные требования, однако, не всегда выполняются в отношении некоторых показателей различия. Например, в случае социальных взаимных предпочтений: Петру нравится Маша, но Маше Петр не нравится. В этом случае симметрия не соблюдена, показатель не может быть метрикой, и такие данные должны быть преобразованы к симметричному виду (разумеется, с потерей информации об асимметрии).

Из критерия симметрии следует, что матрица попарных различий $N \times N$ является симметричной относительно главной диагонали. Из критерия неразличимости идентичных объектов следует, что на главной диагонали этой матрицы различий расположены нули.

Выбор меры различия является органической частью исследования, определяется его процедурой и характером получаемых данных. По процедуре исследования можно выделить три типичные ситуации: испытуемые непосредственно оценивают степень различия (близости) изучаемых объектов; имеются данные об условной или совместной встречаемости событий (объектов); имеется совокупность оценок для множества объектов — используются меры различия профилей оценок. Отметим, что для МШ чаще всего используют непосредственные оценки различий. Остальные меры различий более специфичны для кластерного анализа.

Непосредственная оценка различий

Существует большое количество достаточно хорошо разработанных процедур субъективной оценки различия между объектами: это и прямая оценка различия пары объектов при помощи заданной шкалы (от 5 до 10 градаций)

или графически, и сортировка пар по категориям различия (от идентичных до максимально различных). Основная сложность этих процедур — большое количество пар объектов, предлагаемых испытуемым для сравнения. Для N объектов должно быть оценено количество пар, равное $N(N-1)/2$. Например, если объектов 20, испытуемый должен оценить различия 190 пар объектов.

Самый простой способ организации предъявления всех возможных пар объектов — табличный $N \times N$, когда испытуемый, оценивая различия, заполняет нижний (верхний) треугольник таблицы. Однако при таком предъявлении на оценку различий между объектами влияют не только сами стимулы, но и порядок их предъявления. Речь идет о так называемых пространственных и временных искажениях. *Пространственные искажения* — это влияние порядка следования объектов в каждой паре (второй объект в паре воспринимается иначе, чем первый). *Временные искажения* — это влияние порядка следования пар (пары в начале предъявления воспринимаются иначе, чем в конце ряда). Эти искажения могут быть компенсированы путем случайного упорядочивания порядка следования как объектов в паре, так и самих пар.

При большом числе объектов можно воспользоваться неполным планом, когда каждая пара оценивается не всеми испытуемыми. В неполных планах пары делятся на подмножества, и каждый испытуемый оценивает пары только в одном подмножестве. Каждое из подмножеств оценивается одинаковым числом испытуемых.

При непосредственной оценке различий сложных объектов, например, художественных произведений, людей и т. д., следует иметь в виду принципиальную многомерность таких объектов. Это может иметь своим следствием выделение таких тривиальных критериев-шкал, как рост или пол для людей, толщина или формат для книги и пр. Чтобы избежать таких ситуаций, целесообразно ограничивать контекст сравнения в соответствии с целями исследования — при помощи инструкции или процедуры предъявления пар для сравнения.

ПРИМЕР

В исследовании восприятия студентами учебных курсов, проведенном С. Лященко под нашим руководством[1], студентам предъявлялись пары этих курсов для сравнения. Для ограничения контекста сравнения в инструкции студентам предлагалось представить себе учебный курс, который бы по содержанию во всех отношениях их устраивал («идеал»). Перед сравнением каждой пары объектов каждому испытуемому предлагалось отметить тот из них, который ближе к «идеалу». Ограничение контекста позволило исключить такие тривиальные основания сравнения, как, например, пол преподавателя (именно такой критерий оказался главным на предшествующем этапе исследования, которое проводилось без ограничения контекста сравнения).

---

[1] *Лященко С., Наследов А.* Исследование предпочтений студентами учебных предметов // Психология, акмеология, педагогика – образовательной практике: к 150-летию кафедры педагогики (педагогики и педагогической психологии) и 35-летию ф-та психологии СПбГУ. СПб., 2001.

Результаты социометрии — матрицу взаимоотношений — тоже допустимо рассматривать как вариант прямой оценки различия. Ясно, что такие матрицы — несимметричные, следовательно, получаемые данные требуют предварительной оцифровки и приведения к симметричному виду. Предложенный нами подход продемонстрирован на примере обработки данных социометрии в конце главы 19.

Условные и совместные вероятности

Условные вероятности применяются в исследованиях, где в качестве меры различия объектов (стимулов, событий) x и y выступает вероятность того, что *объект x не встречается при наличии объекта y*. Типичные исследования подобного рода — эксперименты по узнаванию и эксперименты по переходу. В первом случае исследователь предъявляет N стимулов и просит испытуемого каждый раз назвать, какой из N ему был предъявлен. Эксперимент повторяется многократно. По результатам исследования заполняется таблица $N \times N$, строки которой — стимулы, столбцы — ответы. В случае совпадения стимула и ответа к содержимому соответствующей ячейки прибавляется 0, в противном случае — 1. Затем содержимое каждой ячейки делится на количество экспериментов. При таком анализе предполагается, что чем более похожи стимулы, тем чаще они будут путаться. Такие матрицы несимметричны, но приводятся к симметричному виду путем сложения строк и столбцов с одинаковыми номерами. Иначе говоря, исходная матрица складывается с транспонированной.

Аналогичным образом обрабатываются данные в экспериментах по переходу. Строки в матрице $N \times N$ при этом — состояния в начале процесса, столбцы — при окончании (например, профессиональная ориентация до обучения и после). Предполагается, что чем ближе состояния, тем чаще будет переход из одного состояния в другое.

Меры различия профилей для количественных переменных

Профиль — это набор оценок (количественных признаков) объекта. Если объекты — это испытуемые, то профилем могут быть их оценки по каким-либо тестам. Сравниваемыми объектами могут быть и сами признаки, и объекты, оцениваемые испытуемыми, например, по степени предпочтения. Тогда профилем будет совокупность индивидуальных значений соответственно для признака или оцениваемого объекта. Таким образом, меры различия профилей могут применяться к наиболее типичным психологическим данным — типа «объект-признак».

Меры взаимосвязи — самые распространенные и очевидные показатели различия в социальных науках, в том числе в психологии. Иногда, например, в программе STATISTICA, в качестве меры различия используется величина

$1 - r$ (где r — коэффициент корреляции Пирсона). Чем более сходна форма профилей оценок для двух объектов или чем более согласованы изменения индивидуальных оценок для двух признаков, тем более они похожи (выше коэффициент корреляции), а указанная мера различия — ближе к нулю. Следует отметить важную особенность использования коэффициента корреляции в качестве меры различия. Если существует сильная взаимосвязь двух признаков, но она отрицательна (отрицательный коэффициент корреляции), то при анализе эти признаки будут трактоваться как сильно различающиеся. Но если поменять знак одной из переменных, они станут очень близки. Таким образом, коэффициенты корреляции могут являться адекватной мерой различия (близости) только в том случае, если все они положительны (или взяты по абсолютной величине — без учета знака).

Меры расстояния — наиболее специфичные показатели различия именно для профилей, то есть множества P признаков, измеренных для каждого объекта. Различие между двумя объектами, выраженное мерой расстояния, показывает, насколько в среднем различаются совокупности оценок для двух объектов, или насколько в среднем удалены друг от друга профили оценок. Другая трактовка меры расстояния — геометрическая. Это расстояние между двумя точками (объектами) в пространстве, размерность которого равна числу признаков. Значения признаков для объекта трактуются как координаты, задающие его положение в этом пространстве.

Есть два способа задания расстояний, и выбор способа зависит от задачи исследования. Обычно исследователь имеет таблицу (матрицу) данных $N \times P$, где N — строки (испытуемые, объекты), P — столбцы (признаки, оцениваемые объекты). Если исследователя интересуют различия между объектами или испытуемыми (с точки зрения, например, их классификации), то ему необходимо задавать меры различия (расстояния) между строками. Если же ему важны основания классификации признаков, то выбираются меры различия (расстояния) между столбцами.

Существует множество мер расстояния, но наиболее часто используются две: евклидово (или его квадрат) и «метрика города».

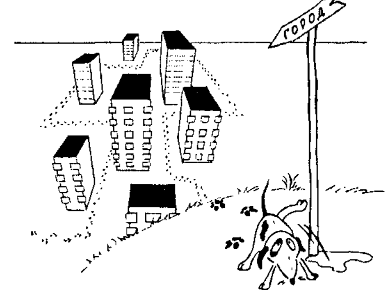
Евклидово расстояние легко представить себе как кратчайшее расстояние (по прямой) между двумя точками (объектами) в пространстве двух, трех и более координат — в соответствии с числом признаков. Наиболее подходит эта метрика для метрических данных (в шкале интервалов или отношений). Если число признаков P, то евклидово расстояние между объектами с номерами i и j выражается формулой:

$$d_{ij} = \left[\sum_{p=1}^{P} (x_{ip} - x_{jp})^2 \right]^{\frac{1}{2}},$$

где x_{ip} и x_{jp} — значения признака p для объектов с номерами i и j соответственно. В пространственном представлении эти значения трактуются как *координаты* объектов i и j по оси p.

«Метрика города» (синонимы — «Манхаттан», «городских кварталов» или *city-block*) — получила свое название по аналогии с длиной пути пешехода,

движущегося от одного объекта к другому в городе. Пешеход вынужден обходить кварталы домов по ломаной линии. Каждый «квартал» при этом — это абсолютная разность между значениями двух объектов по одному из признаков. «Метрика города» больше подходит для неметрических данных (порядковых и, отчасти, бинарных) и вычисляется по формуле:



$$d_{ij} = \sum_{p=1}^{P} \left| x_{ip} - x_{jp} \right|.$$

Важное замечание относительно мер расстояния касается разного масштаба оценок (переменных). Если переменные представлены в разных шкалах, имеют разный масштаб (средние и стандартные отклонения σ), то на расстояние больше будут влиять переменные, имеющие больший разброс (дисперсию). Чтобы уравнять влияние переменных на расстояние между объектами, целесообразно до вычисления расстояний нормировать переменные (делить на стандартное отклонение) или преобразовать их в z-оценки.

Заключая обзор мер различия, следует отметить одно обстоятельство. Наиболее частый вид данных в психологии — это множество признаков, измеренных у множества испытуемых. И наиболее вероятно, что исследователь предпочтет, выбирая многомерное шкалирование или кластерный анализ, опираться на меры различия профилей. В этой ситуации исследователь должен отдавать себе отчет в том, что получаемые показатели различия между объектами не несут никакой иной информации, кроме констатации того, что субъект *A* по измеренным признакам более похож на субъекта *B*, чем на субъекта *C*. А информация о том, по каким признакам различия больше, а по каким — меньше, утрачивается при вычислении расстояний. Таким образом, *одни и те же значения различий между парами объектов могут быть обусловлены разницей в значениях по разным признакам.* Это обстоятельство сильно затрудняет интерпретацию мер расстояния. Иначе дело обстоит, если меры различия профилей используются в отношении данных о субъективных предпочтениях, когда вычисляются меры различия между объектами предпочтения по множеству субъективных оценок (между столбцами — оцениваемыми объектами). Здесь интерпретация достаточно очевидна: объекты будут тем ближе, чем ближе в среднем их располагают субъекты при упорядочивании по степени предпочтения. Вообще говоря, если у исследователя есть выбор при планировании инструментария, для классификации объектов следует предпочесть непосредственные оценки различия, данные о предпочтениях или условные и совместные вероятности.

Меры различия профилей для номинативных переменных

Меры различия для частот — применяются в отношении данных типа «объект-признак», для которых каждый признак представляет собой абсолют-

ную частоту некоторого события для каждого из объектов. В этом случае в качестве мер различия между объектами компьютерные программы предлагают вычислять специфические меры различия для частот (Counts Measures): хи-квадрат (Chi-square) или фи-квадрат (Phi-square). Мера хи-квадрат вычисляется по формуле для χ^2-Пирсона (см. главу 9), а мера фи-квадрат — это величина хи-квадрат, нормализованная путем деления ее на квадратный корень общей суммы частот.

Меры различия для бинарных переменных — применяются, если все переменные набора являются бинарными. Для этого случая в компьютерных программах предусмотрен широкий набор бинарных мер различия (Binary Measures). Например, в программе SPSS предлагается 6 мер для МШ и 27 (!) мер для кластерного анализа. Все эти меры различия основаны на представлении о четырехклеточной таблице сопряженности двух бинарных переменных X и Y:

		Y	
		1	0
X	1	a	b
	0	c	d

где a — количество объектов, для которых и по X и по Y значения равны 1, и т. д. Приведем некоторые меры различий для бинарных переменных.

Квадрат евклидова расстояния (Squared Euclidean distance): $d = b + c$.

Евклидово расстояние (Euclidean distance): $d = \sqrt{b+c}$.

Коэффициент сопряженности φ-Пирсона (Phi 4-points correlation):

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}.$$

Q-коэффициент Юла (Yule's Q): $Q = \frac{ad-bc}{ad+bc}$.

Величина различий (Size difference): $d = \frac{(b-c)^2}{(a+b+c+d)^2}$.

Простой коэффициент совстречаемости (Simple matching): $d = \frac{a+d}{a+b+c+d}$.

То, какая мера предпочтительней, зависит от того, какая роль в исследовании отводится частотам a, $в$, c и d.

НЕМЕТРИЧЕСКАЯ МОДЕЛЬ

Это основной вариант многомерного шкалирования, применяемый в настоящее время. Он лежит в основе всех остальных вариантов метода. Исходные данные для этого метода — матрица размерностью $P \times P$, каждый элемент

которой — мера (оценка) различия между двумя объектами из P. Рассмотрим кратко основные математико-статистические идеи метода, необходимые для его использования на практике.

Многомерное шкалирование и исторически, как новый шаг в математике, и процессуально — как последовательность обработки данных компьютерной программой, начинается с *метрического шкалирования*, предложенного в 50-х годах У. Торгерсоном. В модели Торгерсона вводится жесткое предположение о том, что оценки различия между объектами равны линейному расстоянию между ними в евклидовом пространстве.

Пусть δ_{ij} — имеющаяся в распоряжении исследователя **оценка различия** между объектами i и j. x_{ik} и x_{jk} — координаты этих объектов по оси k, одной из осей искомого пространства размерностью K. **Расстояние между объектами** в искомом пространстве обозначим как d_{ij}. Тогда основное предположение Торгерсона можно выразить формулой:

$$\delta_{ij} = d_{ij} = \left[\sum_{k=1}^{K}(x_{ik} - x_{jk})^2\right]^{\frac{1}{2}}. \qquad (18.1)$$

Торгерсон показал, что при соблюдении этого условия возможен переход от исходной матрицы различий между стимулами к их координатам в пространстве K признаков. Для этого необходимо прежде всего пересчитать исходную матрицу различий в матрицу «скалярных произведений» — путем двойного центрирования, чтобы среднее значение элементов каждой строки и каждого столбца было равно 0. Элементы такой матрицы обозначим как δ^*_{ij}. Тогда, по Торгерсону, справедливо выражение:

$$\delta^*_{ij} = \sum_{k=1}^{K} x_{ik}x_{jk} \text{ или в матричной форме: } \Delta^* = XX',$$

где X — матрица координат стимулов, размерностью $P \times K$.

Это уравнение аналогично главному уравнению факторного анализа, и решается оно относительно X методом главных компонент с заданным числом K.

В современных алгоритмах МШ метод Торгерсона используется на этапе предварительной оценки координат объектов по матрице исходных различий. Далее следует **неметрический этап**, соответствующий неметричности исходных данных. На этом этапе исходят из требования *соответствия рангового порядка расстояний между объектами* в результирующем пространстве *ранговому порядку исходных различий*, то есть, используя принятые обозначения:

$$d_{ij} \leqslant d_{lm} \Rightarrow \delta_{ij} \leqslant \delta_{lm} \text{ для любых } i, j, l, m \text{ — номеров объектов.}$$

Основной мерой выполнения этого требования является специальный показатель, который называется **стресс** — *мера отклонения итоговой конфигурации объектов от исходных оценок различия* в смысле указанного требования рангового соответствия. Иногда дополнительно применяют *коэффициент отчуждения* тоже как меру подгонки неметрической модели к данным о различии.

Не рассматривая подробно вычислительные проблемы многомерного шкалирования, укажем, что его алгоритм направлен на нахождение оценок координат объектов, минимизирующих значение стресса. Построен этот алгоритм как *градиентная процедура*. Первый шаг алгоритма — получение стартовой конфигурации, как правило, методом Торгерсона. На каждом последующем шаге, или итерации, координаты стимулов изменяются в сторону уменьшения значения стресса, вычисленного на предыдущем этапе. Итерации повторяются многократно, до выполнения одного из трех заданных изначально условий (в программе SPSS): достижения минимального значения стресса; достижения минимальной разницы между последним и предыдущим значениями стресса; выполнения максимального заданного числа итераций. Каждое из трех условий задано в программе «по умолчанию», но может изменяться пользователем. Уменьшая пороговые величины стресса и его изменения, увеличивая максимальное число итераций, пользователь может добиться повышения точности окончательного решения. Показателем точности является конечная величина стресса. Наиболее приемлемые величины стресса находятся в диапазоне от 0,05 до 0,2.

Одна из основных проблем, возникающих перед исследователем в МШ — это **проблема размерности *K***. Как и при проведении факторного анализа, в МШ требуется предварительное определение числа шкал. Поэтому от исследователя требуется получить несколько решений в пространствах разной размерности и выбрать из них лучшее. Один из критериев размерности, применяемый для предварительной оценки числа шкал, аналогичен критерию отсеивания Кеттелла в факторном анализе: строится график зависимости стресса от числа шкал по результатам решения в разных размерностях. Истинная размерность соответствует точке перегиба графика после резкого его спада.

Другой критерий числа шкал — абсолютная величина стресса. Если решение одномерно, то приемлемая величина стресса — менее 0,1. Если решение размерностью 2 и выше, то приемлемы значения стресса, меньшие 0,1–0,15. Однако если уровень ошибок измерения или выборки высок, то можно признать решение и с более высокими значениями стресса. Дополнительно вычисляется величина R^2 (RSQ), которая показывает долю дисперсии исходных различий (от единичной), учтенную выделенными шкалами. Чем ближе RSQ к единице, тем полнее данные шкалы воспроизводят исходные различия между объектами.

Окончательный выбор размерности решения определяется на основе критериев интерпретируемости и воспроизводимости, так же, как в факторном анализе. Тем не менее, при размерности 2 и выше, следует избегать решений с величиной стресса выше 0,2. Обычный путь для этого — повышение размерности и исключение объектов.

Результаты применения метода — *таблица координат объектов* в пространстве *K* шкал-признаков, *величины стресса* и *RSQ*, *интерпретация* шкал и взаимного расположения объектов по таблице координат.

ПРИМЕР 18.1

Исследовалась структура представлений студента о многомерных методах, применяемых в психологии. Студенту было предложено сравнить попарно по степени различия пять методов: множественный регрессионный анализ (МРА), дискриминантный анализ (ДА), кластерный анализ (КА), факторный анализ (ФА) и многомерное шкалирование (МШ). При сравнении было предложено использовать 5-балльную шкалу (1 — очень похожи, 5 — совсем разные). Результаты сравнения приведены в табл. 18.4.

Таблица 18.4

Результаты попарного сравнения пяти методов многомерного анализа

Методы	Обозначения	МРА	ДА	КА	ФА	МШ
МРА	MRA	0				
ДА	DA	2	0			
КА	KA	5	2	0		
ФА	FA	2	3	5	0	
МШ	MDS	5	5	3	3	0

Обработка на компьютере

Для обработки воспользуемся данными примера 18.1. Исходные данные (Data Editor) представляют собой нижний треугольник матрицы попарных различий между 5 объектами (табл. 18.4).

1. Выбираем **Analyze > Scale > Multidimensional Scaling (ALSCAL)**...

Примечание. В последних версиях SPSS наряду с вариантом ALSCAL предлагается более современный вариант многомерного шкалирования PROXSCAL. Этот последний вариант, на наш взгляд, действительно более удобен и совершенен. Но поскольку многие пользуются версиями SPSS, в которых программы PROXSCAL еще нет, мы воспользуемся вариантом ALSCAL. Тем более что результаты обработки хоть и различаются, но не существенно. Для тех, кому доступна программа PROXSCAL, не составит большого труда перейти к ней после знакомства с программой ALSCAL.

2. В открывшемся окне диалога переносим из левого в правое верхнее окно (**Variables**) переменные, необходимые для шкалирования (`mra`, `da`, `ka`, `fa`, `mds`). Убеждаемся, что в поле **Distances** (Расстояния) точкой отмечено **Data are distances** (Данные — расстояния), а нажав кнопку **Shape...** (Уточнить) убеждаемся, что матрица данных **Square symmetric** (Симметричная квадратная). Нажимаем **Continue**.

Примечание. Если бы исходные данные были типа «объект-признак», то необходимо было бы воспользоваться опцией **Create distances from data** (Создать расстояния по данным), нажать кнопку **Measure...** (Мера...), задать меру различий и уточнить: между объектами (**Between objects**) или признаками (**Between variables**) вычислять различия. Иногда в качестве исходных данных

применяется несимметричная квадратная матрица различий, например, как результат социометрии. В этом случае указывается соответствующая опция: **Shape... Square asymmetric**.

3. Нажимаем кнопку **Model...** (Модель...) и задаем параметры модели шкалирования. Главным параметром здесь является количество шкал. Обычно следует получить результаты для нескольких шкал и выбрать наилучшее из них — по величинам стресса и по отчетливости интерпретации. В данном случае у нас всего 5 объектов, поэтому вряд ли потребуется более двух шкал. Задаем **Dimensions** (Шкалы) **Minimum:** 2, **Maximum:** 2.

Параметры **Level of measurement** (Уровень измерения) можно не менять и оставить принятые по умолчанию **Ordinal:** (Порядковый:) их изменение практически не меняет результаты. Разве что можно поставить флажок **Untie tied observation** (Корректировать связанные наблюдения) — для устранения влияния связей (повторов) в рангах.

Убеждаемся, что установлено **Conditionality: Matrix** (Условие подгонки: вся матрица).

После задания всех параметров модели нажимаем **Continue**.

4. В основном окне диалога нажимаем **Options** (Опции) для задания параметров обработки и вывода результатов. В появившемся окне диалога внизу в поле **Display** (Выводить) отмечаем флажком **Group plots** (Графики для всей группы) — для графического отображения объектов в координатах шкал. В поле **Criteria** (Критерии) указаны критерии для итераций по подгонке модели: **S-stress convergence:** 0,001 (Величина сходимости s-стресса), **Minimum s-stress value:** 0,005 (Минимальная величина s-стресса), **Maximum iterations:** 30 (Максимальное количество итераций). Эти величины можно не менять.

Примечание. В поле **Criteria** (Критерии) минимальная величина стресса явно занижена: вполне достаточна величина 0,1. Для увеличения точности решения можно увеличить количество итераций и уменьшить величину сходимости стресса (разности стресса на последней и предыдущей итерациях). Но при этом следует внимательно просмотреть по результатам «историю» итераций, так как там могут быть локальные минимумы величины стресса.

После задания всех параметров обработки и вывода результатов нажимаем **Continue**.

Нажимаем OK и получаем результаты.

5. Основные результаты МШ.

А) «История» итераций, величины стресса и RSQ:

```
Iteration history for the 2 dimensional solution
```
(История итераций для 2-шкального решения.)

```
Young's S-stress formula 1 is use
```
(Применена формула s-стресса Юнга.)

```
Iteration      S-stress      Improvement
```
(Итерация) (s-стресс) (Улучшение)

```
    1            .00000
```

```
Iterations stopped because S-stress is less than .005000
```
(Итерации остановлены, поскольку s-стресс меньше, чем 0,005.)

```
Stress values are Kruskal's stress formula 1.
```
(Величина стресса вычислена по формуле 1 Краскала.)

```
For matrix
```
(Для всей матрицы)

```
Stress = .00000 RSQ = 1.00000.
```

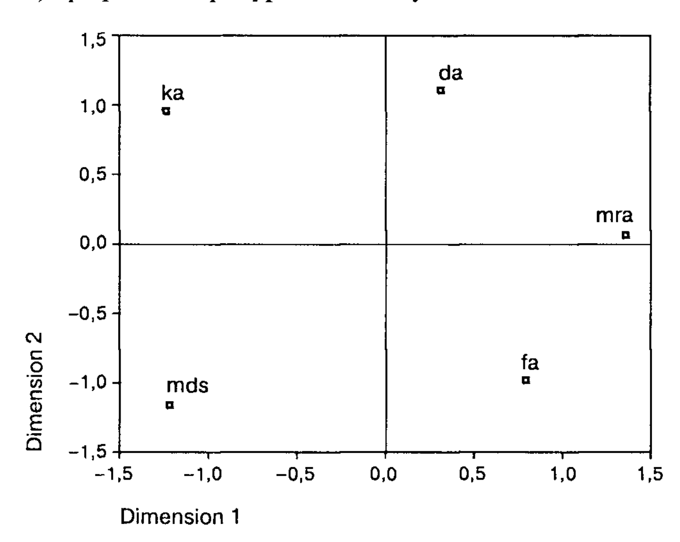
История итераций показывает, что минимальная величина достигнута уже на первом шаге, что на самом деле встречается очень редко. Обычно при большем количестве объектов проблемой является слишком большая величина стресса. Окончательная величина стресса (по формуле 1 Краскала) и величина RSQ свидетельствуют о полном соответствии решения исходным данным.

В) Координаты объектов в осях шкал:

Stimulus Coordinates

Stimulus		Dimension	
		DIM. 1	DIM. 2
1	MRA	1.35	0.04
2	DA	0.44	1.10
3	KA	-1.29	0.95
4	FA	0.72	-0.95
5	MDS	-1.22	-1.13

С) График конфигурации стимулов в осях шкал:



Величина стресса и RSQ свидетельствуют о достаточно точной подгонке конечной конфигурации к исходным данным (расстояния между объектами в итоговом пространстве соответствуют исходным различиям). Следовательно, можно приступать к содержательной интерпретации результатов.

На положительном полюсе первой шкалы располагаются МРА и ФА, на отрицательном — КА и МШ. Промежуточное положение занимает ДА. Таким образом, эта шкала отражает наличие знаний у студента о различии методов по исходным предположениям о структуре данных: ФА, МРА и отчасти ДА исходят из согласованности изменчивости признаков (корреляций), а КА и МШ — из дистантной модели (мер сходства или различия).

На положительном полюсе второй шкалы располагаются ДА и КА — методы классификации, на отрицательном полюсе — ФА и МШ, структурные методы. Следовательно, наличие этой шкалы свидетельствует о сформированности у студента адекватных знаний о назначении многомерных методов. Наглядное представление о субъективной структуре знаний студента дает график координат сравниваемых объектов в пространстве двух шкал.

МОДЕЛЬ ИНДИВИДУАЛЬНЫХ РАЗЛИЧИЙ

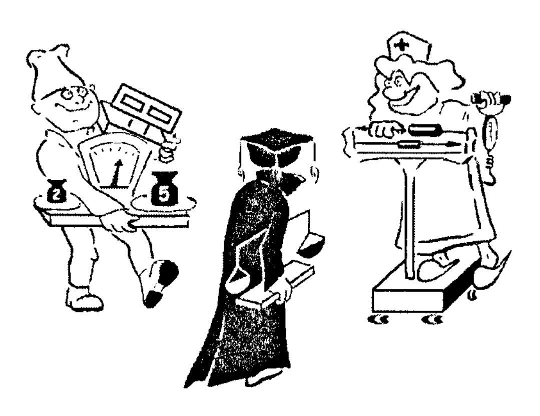
Этот метод — расширение метода неметрического МШ путем включения в основную модель субъективных параметров. Оценки индивидуальных различий дают, наряду с координатами стимулов, количественное *координатное описание субъектов*, аналогичное координатному описанию стимулов, получаемому с помощью неметрического МШ. Исходными данными для МШ индивидуальных различий является N матриц оценок различий P объектов-стимулов, то есть результат оценки попарных различий P объектов каждым из N испытуемых.

Индивидуальная матрица различий для каждого испытуемого может быть получена и в результате вычисления матрицы $P\times P$ по результатам сравнения испытуемым P объектов по ряду признаков. Типичный пример последнего — обработка репертуарных решеток Келли, когда исходными данными являются оценки объектов (например, ролей) по совокупности конструктов. Такой вид анализа легко осуществим при помощи компьютерной программы PROXSCAL, входящей в состав SPSS.

В модели предполагается, что существует общая для всех испытуемых *групповая матрица координат объектов* — X, размерностью $P\times K$ (P — число объектов, K — число координат). При этом для каждого испытуемого имеется индивидуальная *матрица координат* X_i той же размерности, но по содержанию отличающаяся от групповой. Можно представить себе координаты групповой матрицы как совокупность общих (групповых) точек зрения на объекты. Тогда индивидуальная матрица будет отличаться от групповой степенью учета данным испытуемым этих общих точек зрения, иначе — *индивидуальным весом* общих координат.

Таким образом, предполагается, что элементы каждой субъективной матрицы координат отличаются от элементов групповой матрицы *индивидуальными весовыми коэффициентами* — мерами того, насколько учитываются

Каждый субъект может быть охарактеризован набором индивидуальных весов...



данным субъектом общие точки зрения. А каждый субъект может быть охарактеризован набором индивидуальных весов — по одному для каждой координаты (точки зрения). Индивидуальные веса иногда называют весами важности, весами характеристик или масштабными коэффициентами. При прочих равных условиях, при увеличении индивидуального веса данной координаты разность между стимулами по этой координате вносит все больший и больший вклад в оцениваемое различие между этими стимулами.

Изложенное выше можно представить в виде формальных соотношений. Обозначим через x_{ik} координату объекта i по шкале k в общем групповом пространстве объектов. Предполагается, что соответствующий элемент субъективной матрицы координат x_{iks} для испытуемого s связан с элементом групповой матрицы соотношением

$$x_{iks} = x_{ik}w_{ks}$$

или в матричных обозначениях:

$$X_S = XW_s,$$

где w_{ks} — вес координаты k для субъекта s; W_s — матрица субъективных весов для всех K координат, преобразующая групповую матрицу в индивидуальную.

В соответствии с моделью индивидуальных различий исходные данные о субъективном различии объектов i и j субъектом s выражаются формулой

$$\delta_{ijs} = \left[\sum_{k=1}^{K}(x_{iks} - x_{jks})^2\right]^{\frac{1}{2}} = \left[\sum_{k=1}^{K}w_{ks}^2(x_{ik} - x_{jk})^2\right]^{\frac{1}{2}}. \qquad (18.2)$$

Эта формула выражает основное предположение модели МШ индивидуальных различий и связывает исходные данные о *субъективных оценках различий* δ с результатами применения метода: *групповой матрицей координат объектов* X, размерностью $P \times K$, и матрицей субъективных (индивидуальных) весов W, размерностью $N \times K$. При этом матрица X аналогична результатам неметрического МШ, а матрица W содержит N строк — для каждого субъекта и K столбцов — для каждой шкалы-координаты.

Модель индивидуальных различий часто отождествляют с компьютерной программой INDSCALE (индивидуальное шкалирование), предназначенной для нахождения групповой матрицы координат X и индивидуальных весов W по исходным данным о различиях между P объектами, полученным от N субъектов. Эта программа входит, в частности, в состав SPSS. Рассмотрим исходные данные и основные результаты применения этой программы.

Как упоминалось, исходными данными для модели индивидуальных различий являются N матриц субъективных оценок различий между P объектами-стимулами. Мерами соответствия результата исходным данным, или мерами качества решения, являются величины стресса, как и в неметрическом МШ. Дополнительно вычисляются квадраты коэффициентов корреляции (RSQ) между фактическими и оцененными скалярными произведениями. Они показывают степень согласованности расстояний (вычисленных по результирующим координатам стимулов с учетом субъективных весов) с исходными оценками субъективных различий между стимулами. *Общий стресс* и *общий RSQ* — меры соответствия для результирующей групповой матрицы координат стимулов. Они служат, в частности, для оценки числа координат, как и в случае неметрического МШ.

Помимо этого, вычисляются аналогичные меры для каждого субъекта. *Величина стресса и RSQ для субъекта* — это меры соответствия групповой матрицы координат исходным данным для этого субъекта. Чем ниже величина стресса и выше RSQ, тем выше соответствие индивидуальной точки зрения групповой. Если стресс значительно превышает 0,15 и RSQ меньше 0,7, то индивидуальные данные не соответствуют групповым и этого субъекта следует рассматривать отдельно либо выделить в отдельную группу таких же субъектов.

Если величины мер соответствия удовлетворяют исследователя, он может приступить к интерпретации групповой матрицы координат стимулов и матрицы индивидуальных весов. **Матрица координат стимулов** интерпретируется аналогично результатам неметрического МШ: каждая шкала интерпретируется через стимулы, имеющие по этой оси наибольшие абсолютные значения координат. Визуализация матрицы координат стимулов в пространстве двух или трех измерений-шкал позволяет обнаружить содержательно важные группировки объектов.

Матрица индивидуальных весов показывает то, насколько каждый субъект учитывает или разделяет групповые точки зрения при различении объектов. Чем выше для субъекта вес одной из координат, тем существеннее для него соответствующая групповая точка зрения. Матрица индивидуальных весов задает пространство субъектов, каждая ось которого соответствует оси пространства объектов. Каждый субъект в этом пространстве характеризуется вектором из начала координат в точку с координатами, заданными строкой матрицы. Длина каждого вектора прямо пропорциональна степени соответствия индивидуальных данных групповым и пропорциональна RSQ: чем длиннее вектор, тем в большей степени данный субъект учитывает групповые точки зрения. Группировки в пространстве субъектов соответствуют испытуемым со сходными точками зрения.

Отметим, что метод МШ индивидуальных различий предоставляет уникальную возможность сочетания идеографического и нормативного подходов в одном исследовании. Нормативный подход осуществляется путем соотнесения индивидуальных результатов с общими для группы — по индивидуальным субъективным весам общих для группы точек зрения. Одновременно возможно изучение качественного своеобразия каждой индивидуальной точки зрения — по индивидуальной матрице различий.

ПРИМЕР 18.2

В уже упоминавшемся исследовании восприятия студентами учебных предметов[1] каждый из 73 студентов оценивал различия между 14 элементами, в качестве которых выступали пройденные учебные курсы. Полученный массив данных (73 матрицы) обрабатывался при помощи МШ индивидуальных различий. Было получено 3-шкальное решение — достаточно устойчивое и воспроизводимое. Оно оказалось общим для большинства из опрошенных студентов. Для интерпретации шкал далее были проведены структурированные интервью (метод выявления конструктов), материалом для которых являлись учебные дисциплины, поляризованные по шкалам. При проведении интервью учитывалось индивидуальное своеобразие точек зрения опрашиваемого (по результатам шкалирования). Удалось выявить общие конструкты — критерии восприятия студентами учебных курсов. Ими оказались: 1 — «биодетерминизм — социодетерминизм» (в объяснении причин поведения); 2 — «исследование — коррекция» (на чем делается акцент в содержании дисциплины); 3 — «общие — прикладные» (по широте применения или назначения дисциплины). Интересно отметить, что те же конструкты были выявлены и у тех, чьи данные не соответствовали групповым и для которых были составлены индивидуальные поляризации предметов для интервью — в соответствии с результатами шкалирования индивидуальных матриц.

ПРИМЕР 18.3

Изучалась структурированность представлений студентов о разных психологических концепциях. Для этого двум студентам предлагалось сравнить попарно по степени различия концепции пяти ученых: В. Вундта, Э. Титченера, И. М. Сеченова, Э. Торндайка и М. Вертгеймера. В процедуре исследования каждому студенту было предложено оценивать различие концепций в каждой паре из всех возможных сочетаний (всего 5(5 – 1)/2 = 10 пар) по 5-балльной шкале (1 — очень похожи, 5 — совсем не похожи). Результаты оценки различий представлены в табл. 18.5.

Таблица 18.5

Результаты попарного сравнения пяти концепций двумя студентами

Концепции	Обозначения	Студент 1					Студент 2				
		В. Вундт	Э. Титченер	И. М. Сеченов	Э. Торндайк	М. Вертгеймер	В. Вундт	Э. Титченер	И. М. Сеченов	Э. Торндайк	М. Вертгеймер
В. Вундт	v1	0					0				
Э. Титченер	v2	1	0				2	0			
И. М. Сеченов	v3	5	5	0			2	3	0		
Э. Торндайк	v4	5	4	2	0		3	3	3	0	
М. Вертгеймер	v5	4	3	4	3	0	4	4	5	3	0

[1] *Лященко С., Наследов А.* Исследование предпочтений студентами учебных предметов // Психология, акмеология, педагогика – образовательной практике: к 150-летию кафедры педагогики (педагогики и педагогической психологии) и 35-летию ф-та психологии СПбГУ. СПб., 2001.

Обработка на компьютере

Для обработки воспользуемся данными примера 18.3. Исходные данные (**Data Editor**) представляют собой нижние треугольники матриц попарных различий между 5 объектами (табл. 18.4). Для реализации программы INDSCAL необходимо, чтобы матрицы для разных субъектов находились *друг под другом*. В нашем примере вторая матрица расположена под первой.

1. Выбираем **Analyze > Scale > Multidimensional Scaling (ALSCAL)...**

Примечание. Если данные представляют собой оценки объектов по ряду признаков каждым из экспертов (испытуемых), а не матрицы различий, то вместо программы ALSCAL лучше воспользоваться программой PROXSCAL.

2. В открывшемся окне диалога переносим из левого в правое верхнее окно (**Variables**) переменные, необходимые для шкалирования (v1, v2, v3, v4, v5). Убеждаемся, что в поле **Distances** (Расстояния) точкой отмечено **Data are distances** (Данные — расстояния), а нажав кнопку **Shape...** (Уточнить), убеждаемся, что матрица данных **Square symmetric** (Симметричная квадратная). Нажимаем **Continue**.

3. Нажимаем кнопку **Model...** (Модель...) и задаем параметры модели шкалирования. Для данной модели главный параметр — **Scaling model** (Модель шкалирования). Вместо заданной по умолчанию **Euclidean distance** (Евклидово расстояние) задаем **Individual differences Euclidean distance** (Евклидово расстояние индивидуальных различий). В отношении остальных установок руководствуемся теми же соображениями, что и при реализации модели неметрического шкалирования.

Следующим параметром является количество шкал. Обычно следует получить результаты для нескольких шкал и выбрать наилучшее из них — по величинам стресса и по отчетливости интерпретации. В данном случае у нас всего 5 объектов, поэтому вряд ли потребуется более двух шкал. Задаем **Dimensions** (Шкалы) > **Minimum:** 2 (Минимум), **Maximum:** 2 (Максимум). Параметры **Level of measurement** (Уровень измерения) можно не менять и оставить принятые по умолчанию **Ordinal** (Порядковый): их изменение практически не меняет результаты. Разве что можно поставить флажок **Untie tied observation** (Корректировать связанные наблюдения) — для устранения влияния связей (повторов) в рангах.

Убеждаемся, что установлено **Conditionality: Matrix** (Условие подгонки: вся матрица).

После задания всех параметров модели нажимаем **Continue**.

4. В основном окне диалога нажимаем **Options** (Опции) для задания параметров обработки и вывода результатов. В появившемся окне диалога внизу в поле **Display** (Вывод) отмечаем флажком **Group plots** (Графики для всей группы) — для графического отображения объектов в координатах шкал. В поле **Criteria** (Критерий) указаны критерии для итераций по подгонке модели: **S-stress convergence:** 0,001 (Величина сходимости s-стресса), **Minimum s-stress value:** 0,005 (Минимальная величина s-стресса), **Maximum iterations:** 30 (Мак-

симальное количество итераций). Эти величины можно не менять. В отношении этих величин руководствуемся теми же соображениями, что и при реализации модели неметрического шкалирования.

После задания всех параметров обработки и вывода результатов нажимаем **Continue**.

Нажимаем ОК и получаем результаты.

5. Основные результаты МШ индивидуальных различий.

А) «История» итераций:

Iteration (Итерация)	S-stress (s-стресс)	Improvement (Улучшение)
0	.00096	
1	.00074	

```
Iterations stopped because S-stress is less than .005000
```
(Итерации остановлены, поскольку s-стресс меньше, чем 0,005.)

В) Величины стресса и RSQ для каждой матрицы отдельно:

Matrix	Stress	RSQ
1	.000	1.000
2	.001	1.000

Величина стресса и RSQ для всех матриц:

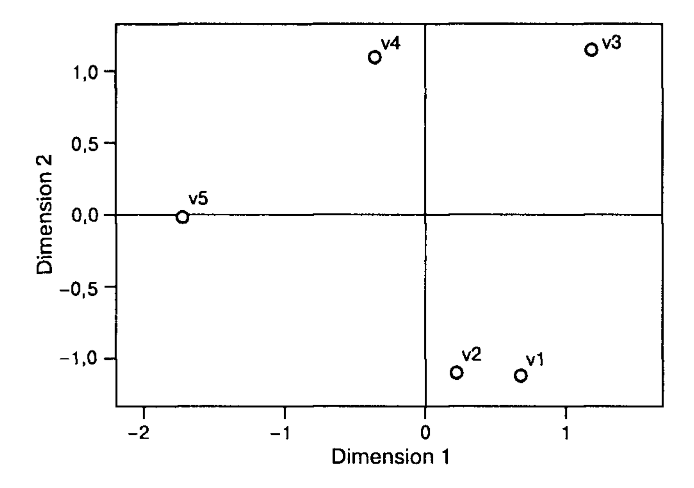
```
Averaged (rms) over matrices
Stress = .00070 RSQ = 1.00000
```

Эти величины свидетельствуют об отличной общей подгонке результатов, в том числе — для каждой матрицы отдельно. Следовательно, можно приступать к интерпретации результатов.

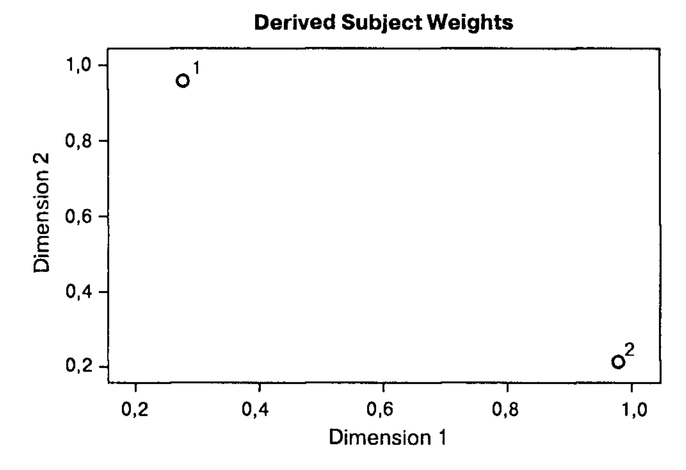
С) Координаты стимулов и субъективные веса для каждой матрицы:

		Dimension (Шкалы)	
Stimulus Number	Stimulus Name	1	2
v1	В. Вундт	0,6701	–1,1265
v2	Э. Титченер	0,2192	–1,1002
v3	И. Сеченов	1,1847	1,1476
v4	Э. Торндайк	–0,3482	1,0969
v5	М. Вертгеймер	–1,7257	–0,0179
Subject Number (Номера субъектов)		Subject Weights (Индивидуальные веса)	
1		0,2798	0,9601
2		0,9768	0,2142

D) График конфигурации стимулов в осях шкал:



E) Конфигурация субъективных весов в осях шкал:



При МШ индивидуальных различий интерпретируются две группы результатов: а) общее для группы испытуемых координатное представление сравниваемых объектов (общие точки зрения); б) субъективные (индивидуальные) веса общих точек зрения для каждого субъекта. На отрицательном полюсе первой шкалы расположена концепция М. Вертгеймера, затем, по мере возрастания значений шкалы: Э. Торндайк, Э. Титченер, В. Вундт, И. Сеченов. Очевидно, что эта шкала отражает временные представления студентов о последовательности появления концепций: чем меньше значения по этой шкале, тем позже появилась концепция. Вторая шкала отражает, скорее, содержательные представления студентов о концепциях: на положительном ее полюсе располагаются концепции, выражающие объективный подход к анализу поведения (И. М. Сеченов, Э. Торндайк); на отрицательном полюсе — интроспективный подход к анализу сознания (В. Вундт, Э. Титченер).

Индивидуальные веса шкал показывают различия испытуемых по тому, насколько каждый из них разделяет общие (групповые) точки зрения. Для студента 1 преимущественное значение имеет содержательная характеристика концепций (шкала 2) и в меньшей степени — последовательность их появления (шкала 1). Студент 2 учитывает в большей степени последовательность появления концепций и почти полностью игнорирует их содержательное своеобразие.

МОДЕЛЬ СУБЪЕКТИВНЫХ ПРЕДПОЧТЕНИЙ

Исходными данными для шкалирования предпочтений является матрица размерностью $P{\times}N$, содержащая N строк — по одной для каждого субъекта, присваивающего номера P объектам по степени предпочтения: от 1 — самому предпочитаемому до P — наименее предпочтительному.

В соответствии с моделью предпочтений каждый субъект характеризуется *идеальным объектом*, а степень предпочтения стимула определяется его отличием от идеала. **Дистанционная модель предпочтений** основана на предположении, что субъекты могут быть охарактеризованы координатами идеальных точек в едином пространстве. Это пространство задается шкалами, которые трактуются как критерии, по которым осуществляются предпочтения. Координата x_{sk} — то значение признака k, которое считает идеальным субъект s. Все значения K признаков определяют набор характеристик идеального объекта. Соответственно, номер или ранг предпочтения определяется как степень отличия данного объекта от идеала — δ_{is}. Чем больше ранг предпочтения, тем меньше нравится объект, то есть тем дальше он от идеала.

Формально неметрическая дистанционная модель предпочтений предполагает выполнение следующих соотношений:

$$\delta_{is} = f_s(d_{is}) = f_s\left[\sum_{k=1}^{K}(x_{ik} - x_{sk})^2\right]^{\frac{1}{2}},\ \delta_{is} < \delta_{js} \Rightarrow d_{is} \leqslant d_{js}\ \text{для всех } (i, j) \text{ субъекта } s. \quad (18.3)$$

Первое соотношение обозначает, что для каждого субъекта есть *своя* монотонная функция f_s, что избавляет от необходимости приписывать субъектам единую шкалу для субъективных предпочтений. Второе соотношение ограничивает координаты объектов x_{ik} и идеальных точек x_{sk} в искомом пространстве так, чтобы сохранить порядковую информацию о соотношении объектов для каждого субъекта. Такой анализ называется *условным по строке* (строки соответствуют субъектам), в отличие от безусловного ограничения (по строкам и столбцам), применяемого в неметрическом шкалировании данных об индивидуальных различиях.

Программа неметрического шкалирования ALSCAL, включенная в состав SPSS, может выполнять и неметрический анализ предпочтений, если задать прямоугольную матрицу (rectangular) с количеством строк (row), соответствующим количеству субъектов, но не менее 4. Дополнительно необходимо задать «условность по строке» (Conditional: Row) в соответствии с требованием выражения 18.3.

Критерии качества координатного представления объектов и правила выбора числа координат при анализе предпочтений те же, что в модели неметрического МШ данных о различиях. Однако матрица координат объектов включает в себя и координаты идеальных точек для каждого субъекта. Иными словами, конечный результат анализа предпочтений — это *групповое пространство признаков* (шкал), в котором наряду с объектами предпочтения размещены **идеальные точки субъектов**. Интерпретация этих результатов аналогична интерпретации результатов анализа различий.

ПРИМЕР 18.4

В упоминавшемся исследовании отношений студентов к учебным предметам (Лященко С., Наследов А., 2001) изучались и их предпочтения. В одной из серий исследования студентам предлагалось упорядочить 14 предметов по степени предпочтения стиля их преподавания. Исходные данные для 73 студентов обрабатывались при помощи многомерного шкалирования предпочтений. Результаты 3-шкального решения использовались для составления структурированных интервью. Таким образом были выделены основные критерии предпочтений учебных курсов с точки зрения стиля их преподавания. Ими оказались: 1 — «академичный — эмоциональный» стиль изложения материала; 2 — «диалогичный — монологичный» характер контакта с аудиторией; 3 — «доступная — сложная» манера изложения материала.

ПРИМЕР 18.5

Исследовались критерии предпочтения студентами различных психологических концепций. Каждому из четырех студентов было предложено ранжировать по степени предпочтения 6 концепций: З. Фрейда, М. Вертгеймера, А. Адлера, Р. Кеттелла, Г. Айзенка и К. Левина (табл. 18.6), присваивая 1 наиболее и 6 наименее предпочитаемой концепции.

Таблица 18.6

Ранги предпочтения студентами шести психологических концепций

Студенты	Концепции					
	З. Фрейд	М. Вертгеймер	А. Адлер	Р. Кеттелл	Г. Айзенк	К. Левин
	v1	v2	v3	v4	v5	v6
1	6	1	5	3	4	2
2	2	6	1	4	3	5
3	3	6	4	2	1	5
4	2	6	3	5	4	1

Обработка на компьютере

Для обработки воспользуемся данными примера 18.5. Исходные данные (**Data Editor**) содержатся в таблице, строки которой соответствуют субъектам, а столбцы — объектам предпочтений (в соответствии с таблицей 18.6).

1. Выбираем **Analyze > Scale > Multidimensional Scaling (ALSCAL)...**

2. В открывшемся окне диалога переносим из левого в правое верхнее окно (**Variables**) переменные, необходимые для шкалирования (v1, v2, v3, v4, v5, v6). Убеждаемся, что в поле **Distances** (Расстояния) точкой отмечено **Data are distances** (Данные — расстояния).

3. Необходимо задать тип матрицы различий. Нажав кнопку **Shape...** (Уточнить) вместо принятой по умолчанию **Square symmetric** (Симметричная квадратная), отмечаем **Rectangular** (Прямоугольная). Указываем число строк, которое должно соответствовать численности экспертов (испытуемых): **Number of rows:** 4 (Количество строк). Нажимаем **Continue**.

4. Нажимаем кнопку **Model...** (Модель...) и задаем параметры модели шкалирования. Для данной модели главный параметр **Conditionality** (Условие подгонки). Вместо заданного по умолчанию **Matrix** (Вся матрица) задаем **Row** (По строке). Убеждаемся, что в поле **Scaling model** (Модель шкалирования) отмечено **Euclidean distance** (Евклидово расстояние). Если в строках часто встречаются одинаковые ранги, то отмечаем флажком **Untie tied observation** (Корректировать связанные наблюдения) — для устранения влияния связей (повторов) в рангах.

Следующим параметром является количество шкал. Обычно следует получить результаты для нескольких шкал и выбрать наилучшее из них — по величинам стресса и по отчетливости интерпретации. В данном случае у нас всего 6 объектов, поэтому вряд ли потребуется более двух шкал. Задаем **Dimensions** (Шкалы) **Minimum:** 2, **Maximum:** 2. После задания всех параметров модели нажимаем **Continue**.

5. В основном окне диалога нажимаем **Options** (Опции) для задания параметров обработки и вывода результатов. В появившемся окне диалога внизу в поле **Display** (Выводить) отмечаем флажком **Group plots** (Графики для всей группы) — для графического отображения объектов в координатах шкал. В поле **Criteria** (Критерии) указаны критерии для итераций по подгонке модели: **S-stress convergence:** 0,001 (Величина сходимости S-стресса), **Minimum s-stress value:** 0,005 (Минимальная величина s-стресса), **Maximum iterations:** 30 (Максимальное количество итераций). Эти величины можно не менять. В отношении этих величин руководствуемся теми же соображениями, что и при реализации модели неметрического шкалирования.

После задания всех параметров обработки и вывода результатов нажимаем **Continue**. Нажимаем OK и получаем результаты.

6. Основные результаты МШ предпочтений.

А) «История» итераций, величины стресса и RSQ:

```
Iteration history for the 2 dimensional solution
```
(История итераций для 2-шкального решения.)

```
Young's S-stress formula 2 is use
```
(Применена формула 2 s-стресса Юнга.)

Iteration (Итерация)	S-stress (s-стресс)	Improvement (Улучшение)
1	0.02095	
2	0.02065	0.0003

```
Iterations stopped because S-stress improvement is less than .001000
```
(Итерации остановлены, поскольку улучшение s-стресса меньше, чем 0,001.)

В) Величины стресса и RSQ для каждой строки отдельно:

```
Stress values are Kruskal's stress formula 2.
```
(Величина стресса вычислена по формуле 2 Краскала.)

Matrix	Stress	RSQ
1	0.000	1.000
2	0.006	1.000
3	0.035	0.999
4	0.046	0.998

Величина стресса и RSQ для всех матриц:

```
For matrix
```
(Для всей матрицы)

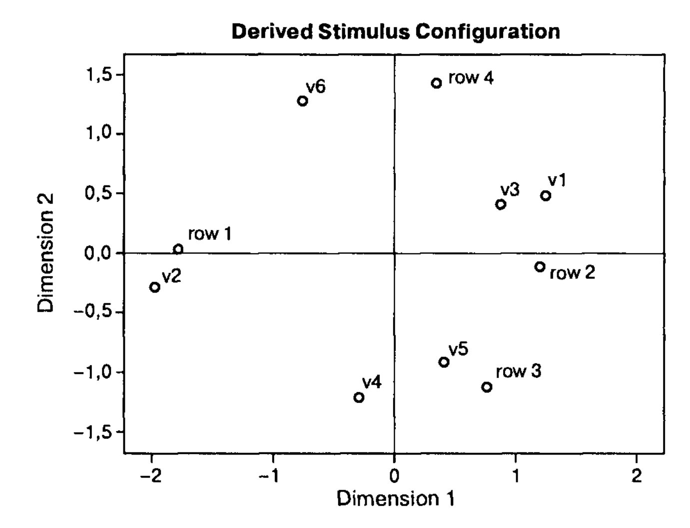
```
 Stress = .029 RSQ = .999.
```

История итераций показывает, что минимальная величина достигнута на втором шаге, что, на самом деле, встречается очень редко. Обычно при большем количестве объектов проблемой является слишком большая величина стресса. Окончательная величина стресса (по формуле 2 Краскала) и величина RSQ свидетельствуют о высоком соответствии исходным данным всего решения. Величины для каждой строки отдельно показывают высокое соответствие исходным данным и результатов для каждого эксперта.

С) Координаты объектов (Column) и идеальных точек (Row) в осях шкал:

Stimulus Number	Stimulus Name	1	2
Column			
v1	З. Фрейд	1.237	0.478
v2	М. Вертгеймер	-1.9721	-0.2875
v3	А. Адлер	0.8707	0.4139
v4	Р. Кеттелл	-0.2983	-1.2141
v5	Г. Айзенк	0.3989	-0.9137
v6	К. Левин	-0.7627	1.2788
Row			
1		-1.7744	0.0344
2		1.1952	-0.1109
3		0.7594	-1.1098
4		0.3464	1.4308

D) График конфигурации объектов и идеальных точек в осях шкал:



Результаты анализа позволяют достаточно определенно интерпретировать основания предпочтений по координатам объектов. Шкала 1 интерпретируется как дихотомия побуждений (З. Фрейд, А. Адлер) и познания (М. Вертгеймер). Шкала 2 противопоставляет концепции, рассматривающие личностные свойства (Р. Кеттелл, Г. Айзенк) и ситуативные условия (К. Левин) в качестве основных причин поведения.

Координаты идеальных точек позволяют идентифицировать индивидуальные субъективные предпочтения. Так, эксперт 1 предпочитает когнитивные концепции, а эксперт 2 — психоанализ.

Глава 19

КЛАСТЕРНЫЙ АНАЛИЗ

НАЗНАЧЕНИЕ

Кластерный анализ решает задачу построения *классификации*, то есть разделения исходного множества объектов на группы (классы, кластеры). При этом предполагается, что у исследователя нет исходных допущений ни о составе классов, ни об их отличии друг от друга. Приступая к кластерному анализу, исследователь располагает лишь информацией о характеристиках (признаках) для объектов, позволяющей судить о сходстве (различии) объектов, либо только данными об их попарном сходстве (различии). В литературе часто встречаются синонимы кластерного анализа: автоматическая классификация, таксономический анализ, анализ образов (без обучения).

Несмотря на то, что кластерный анализ известен относительно давно (впервые изложен Tryon в 1939 году), распространение эта группа методов получила существенно позже, чем другие многомерные методы, такие, как факторный анализ. Лишь после публикации книги «Начала численной таксономии» биологами Р. Сокэл и П. Снит в 1963 году начинают появляться первые исследования с использованием этого метода. Тем не менее, до сих пор в психологии известны лишь единичные случаи удачного применения кластерного анализа, несмотря на его исключительную простоту. Вызывает удивление настойчивость, с которой психологи используют для решения простой задачи классификации (объектов, признаков) такой сложный метод, как факторный анализ. Вместе с тем, как будет показано в этой главе, кластерный анализ не только гораздо проще и нагляднее решает эту задачу, но и имеет несомненное преимущество: результат его применения не связан с потерей даже части исходной информации о различиях объектов или корреляции признаков.

Варианты кластерного анализа — это множество простых вычислительных процедур, используемых для классификации объектов. *Классификация* объектов — это группирование их в классы так, чтобы объекты в каждом классе были более похожи друг на друга, чем на объекты из других классов. Более точно, **кластерный анализ** — это процедура упорядочивания объектов в сравнительно однородные классы на основе попарного сравнения этих объектов по предварительно определенным и измеренным критериям.

Существует множество вариантов кластерного анализа, но наиболее широко используются методы, объединенные общим названием **иерархический кластерный анализ** (*Hierarchical Cluster Analysis*). В дальнейшем под кластерным анализом мы будем подразумевать именно эту группу методов. Рассмотрим основной принцип иерархического кластерного анализа на примере.

ПРИМЕР 19.1

Предположим, 10 студентам предложили оценить проведенное с ними занятие по двум критериям: увлекательность (Pref) и полезность (Use). Для оценки использовалась 10-балльная шкала. Полученные данные (2 переменные для 10 студентов) графически представлены в виде графика двумерного рассеивания (рис. 19.1).

Конечно, классификация объектов по результатам измерения всего двух переменных не требует применения кластерного анализа: группировки и так можно выделить путем визуального анализа. Так, в данном случае наблюдаются четыре группировки: 9, 2, 3 — занятие полезное, но не увлекательное; 1, 10, 8 — занятие увлекательное, но бесполезное; 5, 7 — занятие и полезное и увлекательное; 4, 6 — занятие умеренно увлекательное и умеренно полезное. Даже для трех переменных можно обойтись и без кластерного анализа, так как компьютерные программы позволяют строить трехмерные графики. Но для 4 и более переменных визуальный анализ данных практически невозможен. Тем не менее, общий принцип классификации объектов при помощи кластерного анализа не зависит от количества измеренных признаков, так как непосредственной информацией для этого метода являются различия между классифицируемыми объектами.

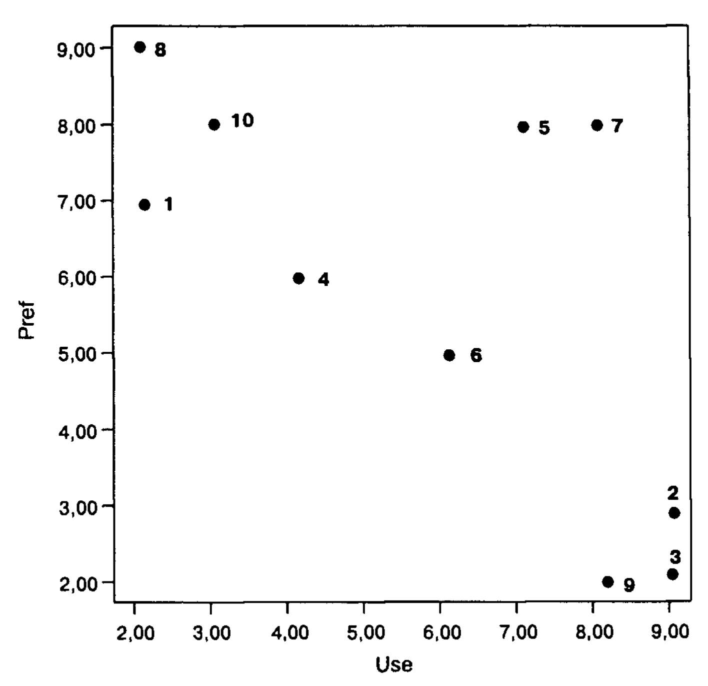


Рис. 19.1. График двумерного рассеивания переменных «увлекательность» (Pref) и «польза» (Use) для 10 студентов

Кластерный анализ объектов, для которых заданы значения количественных признаков начинается с расчета различий для всех пар объектов. Пользователь может выбрать по своему усмотрению меру различия, обзор которых приведен в соответствующем разделе главы 18. В качестве меры различия выбирается расстояние между объектами в P-мерном пространстве признаков, чаще всего — евклидово расстояние или его квадрат. В данном случае $P = 2$ и евклидово расстояние между объектами i и j определяется формулой:

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \ ,$$

где x — это значения одного, а y — другого признака.

На первом шаге кластерного анализа путем перебора всех пар объектов определяется пара (или пары) наиболее близких объектов, которые объединяются в первичные кластеры. Далее на каждом шаге к каждому первичному кластеру присоединяется объект (кластер), который к нему ближе. Этот процесс повторяется до тех пор, пока все объекты не будут объединены в один кластер. Критерий объединения объектов (кластеров) может быть разным и определяется методом кластерного анализа.

Основным результатом применения иерархического кластерного анализа является *дендрограмма* — графическое изображение последовательности объединения объектов в кластеры. Для данного примера дендрограмма приведена на рис. 19.2.

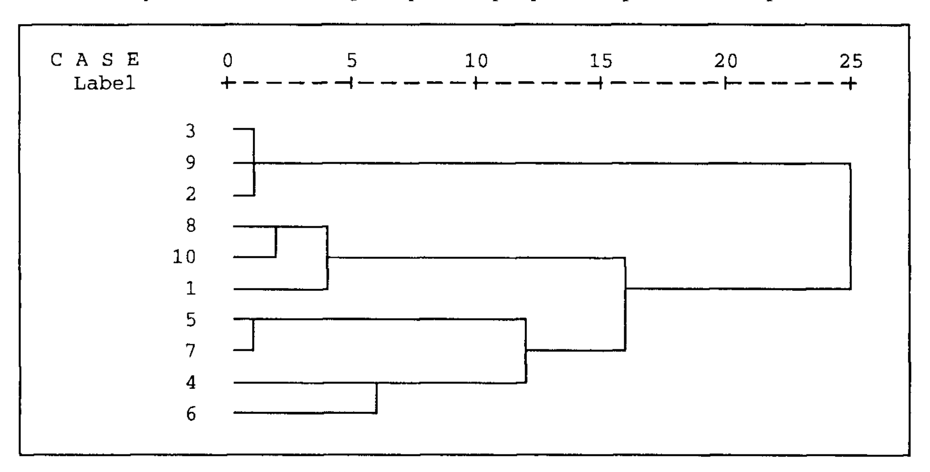


Рис. 19.2. Дендрограмма для 10 студентов (метод средней связи)

На дендрограмме номера объектов следуют по вертикали. По горизонтали отмечены расстояния (в условных единицах), на которых происходит объединение объектов в кластеры. На первых шагах происходит образование кластеров: (3, 9, 2) и (5,7). Далее образуется кластер (8, 10, 1) — расстояния между этими объектами больше, чем между теми, которые были объединены на предыдущих шагах. Следующий кластер — (4, 6). Далее в один кластер объединяются кластеры (5, 7) и (4, 6), и т. д. Процесс заканчивается объединением всех объектов в один кластер. Количество кластеров определяет по дендрограмме сам исследователь. Так, судя по дендрограмме, в данном случае можно выделить три или четыре кластера.

Как видно из примера, кластерный анализ — это комбинаторная процедура, имеющая простой и наглядный результат. Широта возможного примене-

ния кластерного анализа очевидна настолько же, насколько очевиден и его смысл. Классифицирование или разделение исходного множества объектов на различающиеся группы — всегда первый шаг в любой умственной деятельности, предваряющий поиск причин обнаруженных различий.

Можно указать ряд задач, при решении которых кластерный анализ является более эффективным, чем другие многомерные методы:

- разбиение совокупности испытуемых на группы по измеренным признакам с целью дальнейшей проверки причин межгрупповых различий по внешним критериям, например, проверка гипотез о том, проявляются ли типологические различия между испытуемыми по измеренным признакам;
- применение кластерного анализа как значительно более простого и наглядного аналога факторного анализа, когда ставится только задача группировки признаков на основе их корреляции;
- классификация объектов на основе непосредственных оценок различий между ними (например, исследование социальной структуры коллектива по данным социометрии — по выявленным межличностным предпочтениям).

Несмотря на различие целей проведения кластерного анализа, можно выделить общую его последовательность как ряд относительно самостоятельных шагов, играющих существенную роль в прикладном исследовании:

1. *Отбор объектов для кластеризации.* Объектами могут быть, в зависимости от цели исследования: а) испытуемые; б) объекты, которые оцениваются испытуемыми; в) признаки, измеренные на выборке испытуемых.
2. *Определение множества переменных*, по которым будут различаться объекты кластеризации. Для испытуемых — это набор измеренных признаков, для оцениваемых объектов — субъекты оценки, для признаков — испытуемые. Если в качестве исходных данных предполагается использовать результаты попарного сравнения объектов, необходимо четко определить критерии этого сравнения испытуемыми (экспертами).
3. *Определение меры различия* между объектами кластеризации. Это первая проблема, которая является специфичной для методов анализа различий: многомерного шкалирования и кластерного анализа. Применяемые меры различия и требования к ним подробно изложены в главе 18 (раздел «Меры различия»).
4. *Выбор и применение метода классификации* для создания групп сходных объектов. Это вторая и центральная проблема кластерного анализа. Ее весомость связана с тем, что разные методы кластеризации порождают разные группировки для одних и тех же данных. Хотя анализ и заключается в обнаружении структуры, на деле в процессе кластеризации структура привносится в данные, и эта привнесенная структура может не совпадать с реальной.
5. *Проверка достоверности разбиения* на классы.

Последний этап не всегда необходим, например, при выявлении социальной структуры группы. Тем не менее следует помнить, что кластерный анализ

всегда разобьет совокупность объектов на классы, независимо от того, существуют ли они на самом деле. Поэтому бесполезно доказывать существенность разбиения на классы, например, на основании достоверности различий между классами по признакам, включенным в анализ. Обычно проверяют *устойчивость группировки* — на повторной идентичной выборке объектов. *Значимость разбиения* проверяют по внешним критериям — признакам, не вошедшим в анализ.

МЕТОДЫ КЛАСТЕРНОГО АНАЛИЗА

Непосредственными данными для применения любого метода кластеризации является *матрица различий* между всеми парами объектов. Определение или задание меры различия является первым и необходимым шагом кластерного анализа. Поэтому прежде, чем продолжить чтение, убедитесь, что вы уже знакомы с основными мерами различий, с требованиями к ним и со способами их получения (глава 18, раздел «Меры различия»).

Из всего множества методов кластеризации наиболее распространены так называемые *иерархические агломеративные методы*. Название указывает на то, что классификация осуществляется путем последовательного объединения (агломерации) объектов в группы, оказывающиеся в результате иерархически организованными. Эти методы — очень простые комбинаторные процедуры, отличающиеся критерием объединения объектов в кластеры.

Критерий объединения многократно применяется ко всей матрице попарных расстояний между объектами. На первых шагах объединяются наиболее близкие объекты, находящиеся на одном уровне сходства. Затем поочередно присоединяются остальные объекты, пока все они не объединятся в один большой кластер. Результат работы метода представляется графически в виде **дендрограммы** — ветвистого древовидного графика.

Существуют различные методы иерархического кластерного анализа, в частности, в программе SPSS предлагается 7 методов. Каждый метод дает свои результаты кластеризации, но три из них являются наиболее типичными. Поэтому рассмотрим результаты применения этих методов к одним и тем же данным из примера 19.1.



...пока все они не объединятся в один большой кластер

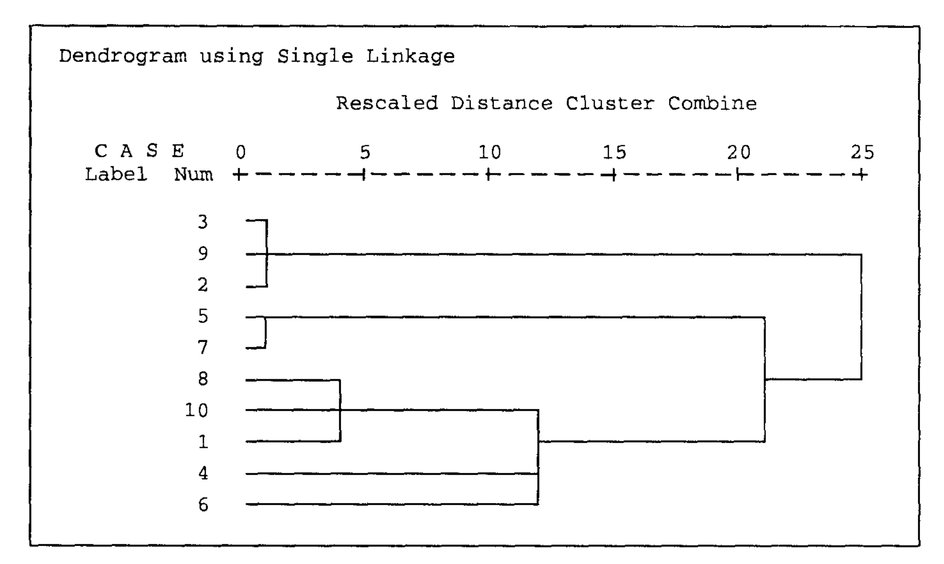


Рис. 19.3. Дендрограмма для 10 студентов (метод одиночной связи)

Метод одиночной связи (*Single Linkage*) — наиболее понятный метод, который часто называют методом *«ближайшего соседа»* (*Nearest Neighbor*). Алгоритм начинается с поиска двух наиболее близких объектов, пара которых образует первичный кластер. Каждый последующий объект присоединяется к тому кластеру, *к одному из объектов которого он ближе*.

На рис. 19.3 приведен результат применения метода. Сопоставляя эту дендрограмму с рис. 19.1, можно заметить, что объект 4 присоединяется к кластеру (8, 10, 1) и на том же расстоянии — к объекту 6 в связи с тем, что расстояние от объекта 4 до объекта 6 такое же, что и до объекта 1. Из рисунка видно, что метод имеет тенденцию к образованию длинных кластеров «цепочного» вида. Таким образом, метод имеет тенденцию образовывать *небольшое число крупных кластеров*. К особенностям метода можно отнести и то, что результаты его применения часто не дают возможности определить, как много кластеров находится в данных.

Метод полной связи (Complete Linkage) часто называют методом *«дальнего соседа»* (*Furthest Neighbor*). Правило объединения этого метода подразумевает, что новый объект присоединяется к тому кластеру, *самый далекий* элемент которого находится *ближе* к новому объекту, чем *самые далекие* элементы других кластеров. Это правило является противоположным предыдущему и более жестким. Поэтому здесь наблюдается тенденция к выделению *большего числа компактных кластеров*, состоящих из наиболее похожих элементов.

Сравним результат применения метода полной связи (рис. 19.4), метода одиночной связи (рис. 19.3) и фактическую конфигурацию объектов (рис. 19.2). Различия в работе методов проявляются прежде всего в отношении объектов 4 и 6. Метод полной связи объединяет их в отдельный кластер и соединяет с кластером (5, 7) раньше, чем с кластером (8, 10, 1) — в отличие от метода одиночной связи. Объект 4 присоединяется сначала к объекту 6, пото-

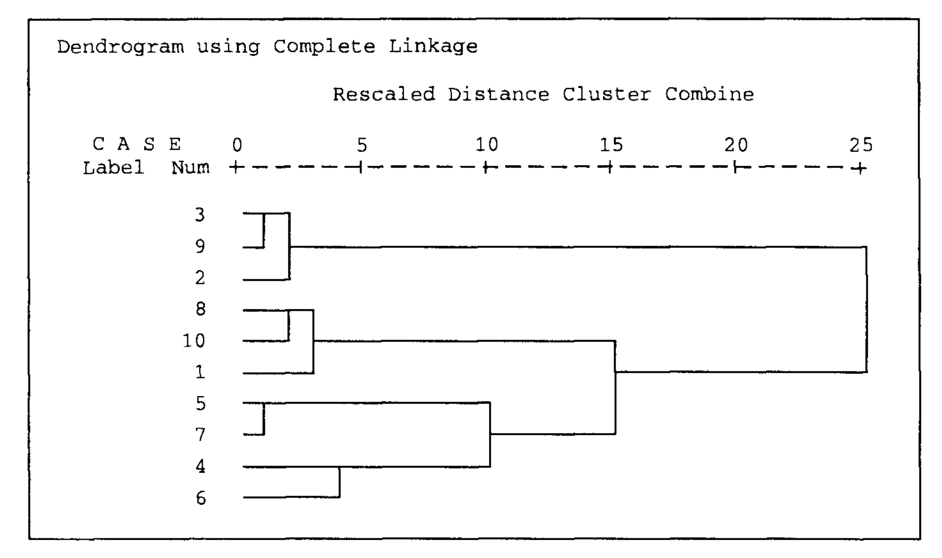


Рис.19.4. Дендрограмма для 10 студентов (метод полной связи)

му что этот последний к нему *ближе, чем самый дальний объект кластера* (8, 10, 1). На этом же основании кластер (4, 6) присоединяется к кластеру (5, 7), потому что самый дальний объект 6 кластера (4, 6) ближе к самому дальнему объекту 7 кластера (5, 7), чем к самому дальнему объекту 8 кластера (8, 10, 1).

Метод средней связи (*Average Linkage*) или *межгрупповой связи* (*Between Groups Linkage*) занимает промежуточное положение относительно крайностей методов одиночной и полной связи. На каждом шаге вычисляется среднее арифметическое расстояние между каждым объектом из одного кластера и каждым объектом из другого кластера. Объект присоединяется к данному кластеру, если это среднее расстояние меньше, чем среднее расстояние до любого другого кластера. По своему принципу этот метод должен давать *более точные результаты классификации*, чем остальные методы. То, что объединение кластеров в методе средней связи происходит при расстоянии большем, чем в методе одиночной связи, но меньшем, чем в методе полной связи, и объясняет промежуточное положение этого метода. Результат применения метода изображен на рис. 19.2. Поскольку объектов в нашем примере немного, результаты применения методов полной и средней связи различаются незначительно.

В реальных исследованиях обычно имеются десятки классифицируемых объектов, и применение каждого из указанных методов дает существенно разные результаты для одних и тех же данных. Опыт и литературные данные свидетельствуют, что наиболее близкий к реальной группировке результат позволяет получить метод средней связи. Но это не означает бесполезность применения двух других методов. Метод одиночной связи «сжимает» пространство, образуя минимально возможное число больших кластеров. Метод полной связи «расширяет» пространство, образуя максимально возможное число компактных кластеров. Каждый из трех методов привносит в реальное

соотношение объектов свою структуру и представляет собой как бы свою точку зрения на реальность. Исследователь, в зависимости от стоящей перед ним задачи, вправе выбрать тот метод, который ему больше подходит.

Численность классов является отдельной проблемой в кластерном анализе. Сложность заключается в том, что не существует формальных критериев, позволяющих определить оптимальное число классов. В конечном итоге это определяется самим исследователем исходя из содержательных соображений. Однако для предварительного определения числа классов исследователь может обратиться к *таблице последовательности агломерации* (*Agglomeration schedule*). Эта таблица позволяет проследить динамику увеличения различий по шагам кластеризации и определить шаг, на котором отмечается резкое возрастание различий. Оптимальному числу классов соответствует разность между числом объектов и порядкового номера шага, на котором обнаружен перепад различий. Более подробно порядок оценки численности классов рассмотрен на примере компьютерной обработки.

Обработка на компьютере: кластерный анализ объектов

Воспользуемся для обработки на компьютере данными примера 19.1. Исходные данные (Data Editor) представляют собой два столбца (переменные Use и Pref) и 10 строк.

1. Выбираем **Analyze** > **Classify** (Классификация) > **Hierarchical Cluster...** (Иерархический кластерный).

2. В открывшемся окне диалога переносим из левого в правое верхнее окно (**Variables**) переменные, необходимые для анализа (`Pref`, `Use`). Убеждаемся, что в поле **Cluster** точка установлена на **Cases** (Объекты), а не на **Variables** (Переменные) — эта установка задает то, что будет подлежать классификации: объекты или переменные. Убеждаемся, что в поле **Display** (Выводить) флажки установлены на **Statistics** (Статистики), **Plots** (Графики).

3. Нажимаем клавишу **Statistics...** (Статистики...) и убеждаемся, что установлен флажок на **Agglomeration schedule** (Последовательность агломерации). При необходимости можно было бы отметить и **Proximity matrix** (Матрица расстояний) для ее вывода, но мы этого не делаем. Нажимаем **Continue** (Продолжить).

4. Нажимаем клавишу **Plots...** (Графики...). Отмечаем флажком **Dendrogram** (Дендрограмма). Здесь же можно выбрать ориентацию дендрограммы: вертикальную (**Vertical**) или горизонтальную (**Horizontal**), оставляем установленную по умолчанию вертикальную ориентацию. Нажимаем **Continue**.

5. Нажимаем **Method...** (Метод...), и открывается окно главных установок кластерного анализа. В этом окне четыре поля установок метода кластеризации: **Cluster Method** (Метод кластеризации), **Measure** (Меры различия), **Transform Values** (Преобразование значений признаков), **Transform Measures** (Преобразование мер различия). В поле **Cluster Method** (Метод кластериза-

ции) оставляем принятый по умолчанию **Between-groups linkage** (Метод средней связи). В поле **Measure** (Меры различия) выбираем **Interval data: Euclidean distance** (Интервальные данные: Евклидово расстояние). Остальные установки оставляем принятыми по умолчанию. Нажимаем **Continue**. Нажимаем ОК и получаем результаты.

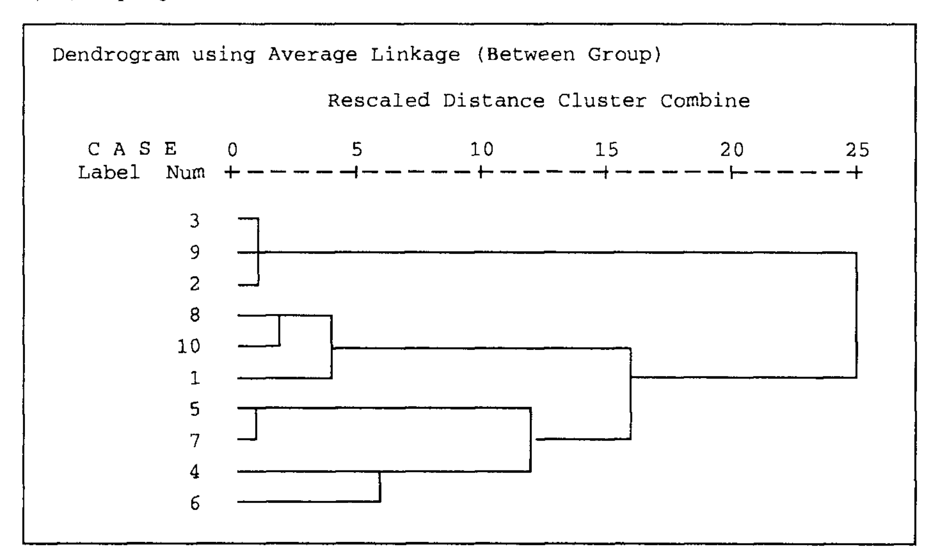
6. Основные результаты кластерного анализа.

А) Таблица последовательности агломерации:

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	3	9	1.000	0	0	3
2	5	7	1.000	0	0	7
3	2	3	1.207	0	1	9
4	8	10	1.414	0	0	5
5	1	8	1.707	0	4	8
6	4	6	2.236	0	0	7
7	4	5	3.711	6	2	8
8	1	4	4.484	5	7	9
9	1	2	6.726	8	3	0

В) Дендрограмма:



Помимо дендрограммы, очень важна информация, содержащаяся в таблице последовательности агломерации. В этой таблице вторая колонка Cluster Combined (Объединенные кластеры) содержит первый (Cluster 1) и второй

(Cluster 2) столбцы, которые соответствуют номерам кластеров, объединяемых на данном шаге. После объединения кластеру присваивается номер, соответствующий номеру в колонке Cluster 1. Так, на первом шаге объединяются объекты 3 и 9, кластеру присваивается номер 3, далее этот кластер на шаге 3 объединяется с элементом 2, новому кластеру присваивается номер 2 и т. д. Следующая колонка Coefficients (Коэффициент) содержит значение расстояния между кластерами, которые объединяются на данном шаге. Колонка Stage Cluster First Appears (Предыдущий шаг, на котором появлялся кластер) показывает, на каком шаге *до этого* появлялся первый и второй из объединяемых кластеров. Последняя колонка Next Stage (Следующий шаг) показывает, на каком шаге снова появится кластер, образованный на этом шаге.

Попытаемся оценить оптимальное число классов по таблице последовательности агломерации. Видно, что первый резкий скачок расстояния между кластерами наблюдается при переходе от 6 к 7 шагу. Следовательно, наиболее оптимальное количество кластеров — то, которое получено на 6 или 7 шаге. Это количество равно численности объектов минус номер шага, то есть 10 – 6 (7) = 4 (3) – 4 или 3 кластера. Выбор того или иного решения будет зависеть уже от содержательных соображений. Так, в данном случае, если обратиться к рис. 19.1, то целесообразно выделять 4 кластера, то есть отделять кластеры (4, 6) — умеренные оценки и (5, 7) — высокие оценки увлекательности и полезности занятия.

КЛАСТЕРНЫЙ И ФАКТОРНЫЙ АНАЛИЗ

Как отмечалось ранее, кластерный анализ можно применять в ходе корреляционного анализа — для исследования взаимосвязей множества переменных, как существенно более простой и наглядный аналог факторного анализа. В этом смысле представляет интерес соотнесение факторного и кластерного анализа.

Факторный анализ (глава 16), как известно, позволяет выделить факторы, которые интерпретируются как латентные причины взаимосвязи групп переменных. При этом каждый фактор идентифицируется (интерпретируется) через группу переменных, которые теснее связаны друг с другом, чем с другими переменными. Напомним, что кластерный анализ тоже направлен на выявление групп, в состав которых входят объекты, более сходные друг с другом, чем с представителями других групп. При этом, конечно же, кластерный анализ имеет совершенно иную природу, нежели факторный анализ. Но если в качестве объектов классификации определить *переменные*, а в качестве мер их различия (близости) — *корреляции*, то кластерный анализ позволит получить тот же результат, что и факторный анализ. Имеется в виду доступная интерпретации структура взаимосвязей множества переменных.

Важно отметить два существенных *ограничения факторного анализа*. Во-первых, факторный анализ неизбежно сопровождается потерей исходной

информации о связях между переменными. И эта потеря часто весьма ощутима: от 30 до 50%. Во-вторых, из требования «простой структуры» следует, что ценность представляет решение, когда группы переменных, которые соответствуют разным факторам, не должны заметно коррелировать друг с другом. И чем теснее эти группы связаны, тем хуже факторная структура, тем труднее факторы поддаются интерпретации. Не говоря уже о случаях иерархической соподчиненности групп.

Кластерный анализ корреляций лишен указанных недостатков. Во-первых, классификация при помощи кластерного анализа по определению отражает всю исходную информацию о различиях (связях в данном случае). Во-вторых, он не только допускает, но и отражает степень связанности разных кластеров, включая случаи соподчиненности (иерархичности) кластеров.

Таким образом, кластерный анализ является не только более простой и наглядной альтернативой факторного анализа. В указанных отношениях он имеет явные преимущества, которые целесообразно использовать, по крайней мере, до попытки применения факторного анализа. Как начальный этап исследования корреляций, кластерный анализ позволит избавиться от несгруппированных переменных и выявить иерархические кластеры, к которым факторный анализ не чувствителен. Вполне вероятно, что после кластерного анализа отпадет и сама необходимость в проведении факторного анализа. Исключение составляют случаи применения факторного анализа по его прямому назначению — для перехода к факторам как к новым интегральным переменным.

Применяя кластерный анализ для исследования структуры корреляций, необходимо помнить о двух обстоятельствах. Во-первых, *корреляция является мерой сходства*, а не различия — ее величина возрастает (до 1) при увеличении сходства двух переменных. Во-вторых, отрицательные величины корреляции так же свидетельствуют о сходстве переменных, как и положительные, то есть для классификации необходимо использовать только положительные корреляции (их абсолютные значения).

ПРИМЕР 19.2

В конце главы 16 был рассмотрен пример применения факторного анализа в психосемантическом исследовании. Напомним, что в результате многоэтапной обработки была получена 3-факторная структура для 10 переменных — шкал семантического дифференциала (табл. 19.1). В качестве исходных данных выступали 10 переменных, измеренных для 86 объектов. Для сравнения применим кластерный анализ в отношении тех же данных для классификации 10 переменных, используя в качестве меры различия *абсолютное* значение коэффициента корреляции Пирсона.

Таблица 19.1

Факторная структура 10 шкал СД

Переменные		Факторные нагрузки (после вращения)		
№	Название	1	2	3
v1	Разговорчивый — Молчаливый	0,09	–0,09	–0,82
v2	Безответственный — Добросовестный	0,02	0,58	0,04

Окончание табл. 19.1

Переменные		Факторные нагрузки (после вращения)		
v3	Замкнутый — Открытый	0,23	0,20	0,76
v4	Зависимый — Независимый	0,70	0,21	–0,03
v5	Деятельный — Пассивный	0,03	–0,63	–0,39
v6	Вялый — Энергичный	0,20	0,65	0,45
v7	Расслабленный — Напряженный	–0,94	–0,07	–0,11
v8	Суетливый — Спокойный	0,81	–0,11	0,01
v9	Несамостоятельный — Самостоятельный	0,28	0,80	0,01
v10	Раздражительный — Невозмутимый	0,66	0,32	0,13

Обработка на компьютере: кластерный анализ корреляций

В качестве исходных данных (**Data Editor**) воспользуемся теми же данными, на основании которых были получены результаты факторного анализа (табл. 19.1), то есть 10 признаков для 86 объектов.

1. Выбираем **Analyze > Classify** (Классификация) > **Hierarchical Cluster...** (Иерархический кластерный).

2. В открывшемся окне диалога переносим из левого в правое верхнее окно (**Variables**) переменные, необходимые для анализа (v1, v2, ..., v10). В поле **Cluster** устанавливаем точку на **Variables** (Переменные), а не на **Cases** (Объекты), эта установка задает то, что подлежать классификации будут переменные. Убеждаемся, что в поле **Display** (Выводить) флажки установлены на **Statistics** (Статистики), **Plots** (Графики).

3. Нажимаем **Method...** (Метод...), и открывается окно главных установок кластерного анализа. В поле **Measure** (Меры различия) выбираем **Interval data: Pearson correlation** (Интервальные данные: Корреляция Пирсона). В поле **Transform Measures** (Преобразование мер различия) устанавливаем флажок **Absolute values** (Абсолютные значения). В поле **Cluster Method** (Метод кластеризации) оставляем принятый по умолчанию Метод средней связи (**Between-groups linkage**). Нажимаем **Continue**.

4. Нажимаем клавишу **Statistics...** (Статистики) и убеждаемся, что установлен флажок на **Agglomeration schedule** (Последовательность агломерации). При необходимости можно было бы отметить и **Proximity matrix** (Матрица расстояний) для ее вывода (будет выведена матрица корреляций), но мы этого не делаем. Нажимаем **Continue** (Продолжить).

5. Нажимаем клавишу **Plots...** (Графики). Отмечаем флажком **Dendrogram** (Дендрограмма). Здесь же можно выбрать ориентацию дендрограммы: вертикальную (**Vertical**) или горизонтальную (**Horizontal**), оставляем установленную по умолчанию вертикальную ориентацию. Нажимаем **Continue**. Нажимаем OK и получаем результаты.

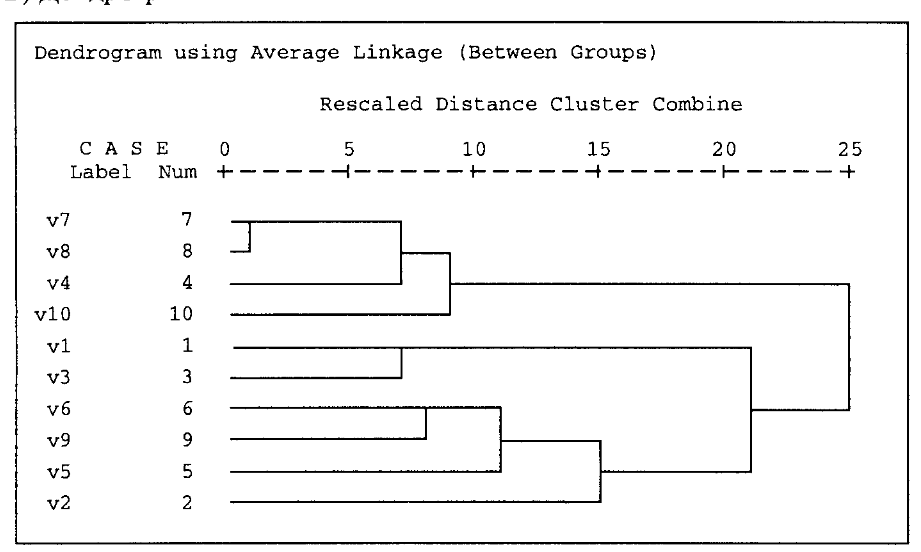
6. Основные результаты кластерного анализа.

А) Таблица последовательности агломерации:

Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	7	8	.758	0	0	2
2	4	7	.620	0	1	5
3	1	3	.607	0	0	8
4	6	9	.594	0	0	6
5	4	10	.557	2	0	9
6	5	6	.524	0	4	7
7	2	5	.422	0	6	8
8	1	2	.280	3	7	9
9	1	4	.186	8	5	0

В) Дендрограмма:



По таблице последовательности агломерации резкое уменьшение величины корреляции между кластерами наблюдается после шага 7, на котором объединяются кластеры 2 и 5 (переменная 2 присоединяется к кластеру 6, 9, 5). Следовательно, оптимальное число кластеров равно 3. Этот результат подтверждает дендрограмма. Состав кластеров точно соответствует факторной структуре (табл. 19.1). В дополнение к результатам факторного анализа можно добавить, что группы переменных, которые соответствуют факторам 2 и 3, заметно коррелируют друг с другом ($r = 0{,}280$, $p < 0{,}05$), то есть фактор «добросовестность» положительно связан с фактором «экстраверсия».

Таким образом, если бы задача исследования была ограничена обнаружением и интерпретацией групп переменных, то кластерный анализ позволил бы ее решить быстрее, проще и в некотором смысле полнее.

КЛАСТЕРНЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ СОЦИОМЕТРИИ

Социометрия — широко применяемый метод для изучения социальной структуры группы по результатам положительных и отрицательных выборов ее членами друг друга. Как правило, каждому представителю группы предлагается по указанному экспериментатором критерию выбрать несколько наиболее (положительный выбор) и наименее (отрицательный выбор) подходящих представителей данной группы. Число выборов задается исследователем в зависимости от размера группы, обычно — от 2 до 5. Результаты выборов фиксируются в таблице, социоматрице, строки которой соответствуют тем, кто выбирает, а столбцы — кого выбирают. Наиболее сложная и трудоемкая задача социометрии — выделение группировок в исходной группе, решение которой при «ручной» обработке находится на грани возможностей, если численность группы — несколько десятков человек.

Кластерный анализ социометрической матрицы не только прост в исполнении, но и позволяет выделить разные «точки зрения» на социальную структуру группы.

В табл. 19.2 приведены результаты социометрии в группе из 12 человек. Плюсы — положительные выборы, минусы — отрицательные, пропуски — отсутствие выборов.

Таблица 19.2

Симметричная матрица для 12 испытуемых

N	1	2	3	4	5	6	7	8	9	10	11	12
1	0	+	+	+	–	–			+		+	
2	+	0								–	+	
3	+		0	+						+		
4				0			+	+	+	–	–	
5	–				0	+	–	+				–
6	–				+	0	+	–				–
7	–					+	0	+	–			+
8						–	+	0			+	+
9	–			+	–				0	+	+	
10	+	+			–	–		–	+	0	+	
11	–	+								+	0	–
12	–				–	–	+	+			+	0

Первая проблема — задание метрики взаимоотношений. В социометрии субъект далеко не всегда выбирает того, кто выбрал его, следовательно, социоматрица не симметрична относительно главной диагонали, и ее элементы не

являются мерами различия. В то же время исходной для кластерного анализа должна быть матрица, симметричная относительно главной диагонали, каждый элемент которой отражает попарные различия (расстояния) между субъектами. Следовательно, необходимо оцифровать исходную социоматрицу так, чтобы ее можно было представить в симметричном виде. При этом величина различий должна быть мерой симпатии-антипатии между членами группы: чем выше симпатия, тем меньше эта величина, а чем выше антипатия, те больше величина различий.

В соответствии с тремя типами одностороннего отношения («да» — положительный выбор, «нет» - отрицательный выбор, «пусто» — отсутствие выбора) можно определить шесть возможных типов взаимоотношений, каждому из которых необходимо присвоить свое числовое значение, отражающее социальную дистанцию. Минимальной дистанции («да–да» — положительный взаимный выбор) должно соответствовать минимальное число, максимальной («нет–нет» — отрицательный взаимный выбор) — максимальное число. Исходя из этого принципа, можно подобрать числа для исходных выборов. Положительному выбору («да») присвоим значение 1, нейтральному отношению («пусто») — 2, отрицательному выбору («нет») — 4. Другой вариант, соответственно, 1–3–4. Вариант 1–2–3 недопустим, так как будут неразличимы варианты отношений «да–нет» и «пусто–пусто». Итак, на данном этапе мы заменяем варианты выборов: «+» на 1, «пусто» на 2, «–» на 4. Диагональные элементы оставляем нулями (расстояния между идентичными элементами).

Следующий шаг — приведение исходной матрицы к симметричному виду — производится путем сложения строк и столбцов с одинаковыми номерами (иначе говоря, исходная матрица складывается с самой собой транспонированной). Полученная симметричная матрица (табл. 19.3) отражает все попарные взаимоотношения между членами группы: «да–да» — 2, «да–пусто» — 3, «пусто–пусто» — 4, «да–нет» — 5, «пусто–нет» — 6, «нет–нет» — 8.

Таблица 19.3

Симметричная социометрическая матрица
(после оцифровки данных табл. 19.2)

N	1	2	3	4	5	6	7	8	9	10	11	12
1	0	2	2	3	8	8	6	4	5	3	5	6
2	2	0	4	4	4	4	4	4	4	5	2	4
3	2	4	0	3	4	4	4	4	4	3	4	4
4	3	4	3	0	4	4	3	3	2	6	6	4
5	8	4	4	4	0	2	6	3	6	6	4	8
6	8	4	4	4	2	0	2	8	4	6	4	8
7	6	4	4	3	6	2	0	2	6	4	4	2
8	4	4	4	3	3	8	2	0	4	6	3	2
9	5	4	4	2	6	4	6	4	0	2	3	4
10	3	5	3	6	6	6	4	6	2	0	2	4
11	5	2	4	6	4	4	4	3	3	2	0	5
12	6	4	4	4	8	8	2	2	4	4	5	0

В отношении полученной симметричной матрицы необходимо сделать два замечания. Во-первых, на этом этапе мы теряем часть исходной информации — о направлении выбора. Например, расстояние между субъектами 1 и 4 равно 3. Это свидетельствует об одностороннем положительном выборе в паре, однако симметричная социоматрица не дает возможности узнать, кто в паре выбирает, а кто остается нейтральным. Во-вторых, принятая в данном случае оцифровка (1–3–4) приводит к тому, что расстояние в паре с амбивалентными отношениями («да–нет» = 5) становится меньше, чем взаимно нейтральные отношения («пусто–пусто» = 6). При оцифровке 1–2–4 соотношение было бы обратным. В зависимости от того, какие отношения исследователь считает более близкими — амбивалентные или односторонне отрицательные, выбирается и вариант оцифровки.

Рассмотрим теперь результаты применения к этой матрице **метода одиночной связи** (рис. 19.5). На уровне взаимного положительного выбора (расстояние равно 2) образовано два кластера. Отношение между кластерами — на уровне одностороннего положительного выбора.

Каждый субъект присоединяется к одному из двух кластеров на основании наличия взаимного положительного выбора между ним и одним из членов этого кластера. При этом не учитываются его отношения с другими членами этого кластера, в том числе и взаимно отрицательные. Например, субъект 12 присоединен ко второму кластеру на основании взаимного положительного выбора с субъектом 7 или 8, несмотря на то что его отношения с остальными членами этого кластера (5 и 6) находятся на уровне взаимного неприятия. Иначе говоря, при использовании этого метода в расчет принимаются только наиболее близкие отношения. Новый кандидат на включение присоединяется к той группировке (кластеру), один из членов которой находится в самых близких с ним отношениях по сравнению с членами других группиро-

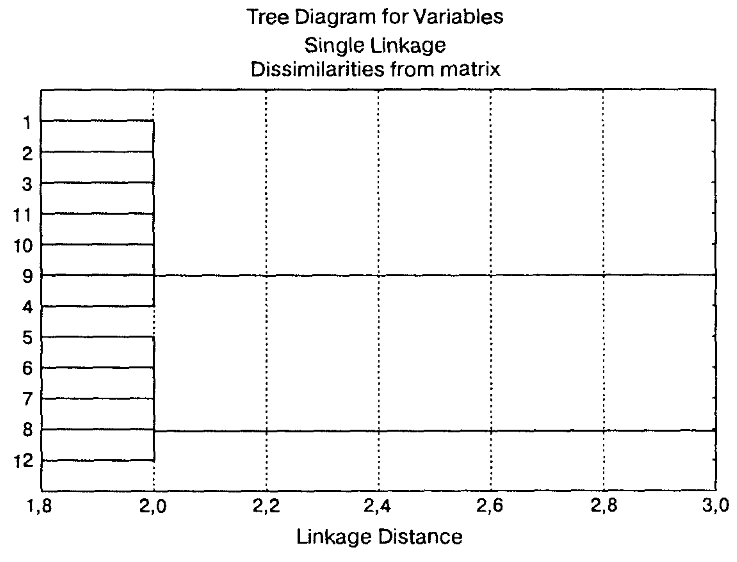


Рис. 19.5. Метод одиночной связи в применении к данным социометрии

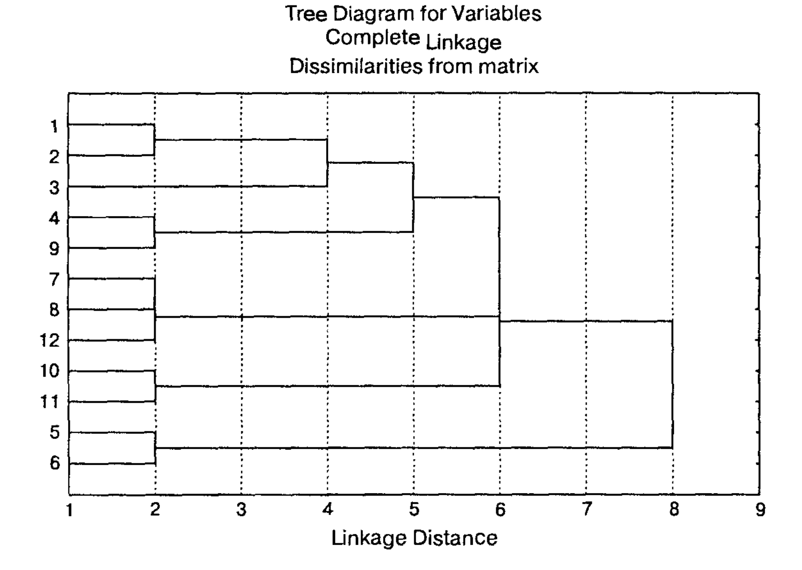


Рис. 19.6. Метод полной связи в применении к данным социометрии

вок. Отношения между двумя выделенными группировками трактуются так же: это наилучшие взаимоотношения между одним из членов одной группировки и одним из членов другой, в данном случае они соответствуют расстоянию 3 — одностороннему положительному выбору.

Метод полной связи демонстрирует другой подход к анализу той же матрицы (рис. 19.6). На первом шаге, на уровне взаимного положительного выбора, образуется 6 кластеров: четыре пары, одна тройка и один тривиальный кластер, состоящий из одного члена. Далее, при установлении отношений (расстояния) между новым кандидатом на включение в кластер и этим кластером, из всех расстояний между ним и членами этого кластера выбирается наибольшее. Поэтому на втором шаге к группировке 1–2 присоединяется субъект 3 на уровне взаимно нейтральных отношений. На следующем шаге к группировке 1–2–3 присоединяется группировка 4–9. Отношения между ними определяются как амбивалентные (расстояние 5), в соответствии с тем, что это наихудшие отношения между членами этих группировок. Эти отношения, однако, лучше, чем были бы при объединении любой из этих двух группировок с одной из остальных. Точно так же происходит дальнейшее объединение кластеров — на основании наихудших отношений в объединяемых группировках, которые, однако, должны быть лучше, чем сложились бы при объединении с другими группировками. Последние две группировки объединяются на уровне взаимно отрицательных отношений. Это означает, что такие отношения не встречаются отдельно в каждой из группировок, но существуют между некоторыми членами одной группировки и некоторыми — другой, то есть появляются при их объединении.

Таким образом, результаты применения разных методов кластеризации соответствуют разным точкам зрения на одну и ту же структуру взаимоотно-

шений. С точки зрения члена группы, результаты применения метода одиночной связи соответствуют принципу «друг моего друга — свой, хоть и враг мне». Результаты применения метода полной связи больше соответствуют другой точке зрения: «друг моего врага — чужой, хоть и друг мне». При этом «друг» и «враг» отличаются расстоянием, а «свой» и «чужой» — принадлежностью к группировке (кластеру). Иначе говоря, в отношении одной и той же социальной структуры метод одиночной связи соответствует точке зрения оптимиста, а метод полной связи — пессимиста.

Обработка на компьютере: кластерный анализ различий

Укажем последовательность шагов для обработки данных социометрии. Симметричную матрицу различий (табл. 19.3) можно получить при помощи программы Excel. Для этого сначала необходимо набрать социометрическую матрицу (табл. 19.2). Затем при помощи операции «заменить» оцифровать всю матрицу, заменяя символы и пустые клетки соответствующими цифрами. После этого при помощи копирования и специальной вставки (отметить «транспонировать») поместить рядом с исходной оцифрованной матрицей ее транспонированную копию. Симметричная матрица различий получается как третья матрица путем поэлементного сложения исходной оцифрованной матрицы с ее транспонированной копией.

При обработке матрицы различий при помощи статистических программ (SPSS или STATISTICA) возникает неожиданная проблема. Дело в том, что ни в той, ни в другой программе *не предусмотрен непосредственный ввод матриц различий для обработки при помощи кластерного анализа* (в SPSS модули многомерного шкалирования позволяют обрабатывать подобного рода данные, а модули кластерного анализа — нет). Требуется ввод таких матриц в особом матричном формате (в SPSS — еще и с использованием специальной коррекции программ обработки). В связи с тем, что это ограничение легче «обойти» в программе STATISTICA, приведем последовательность обработки симметричной матрицы различий в среде именно этой программы.

1. Подготавливаем таблицу исходных данных (Spreadsheet) требуемой размерности, в данном случае — 12×12. Путем копирования и вставки переносим матрицу различий из таблицы Excel в таблицу Data: Spreadsheet (STATISTICA).

2. Открываем диалог метода кластерного анализа: **Statistics... > Cluster Analysis**. Не меняя установок по умолчанию (программа воспринимает данные как Row Data — типа «объект-признак»), выбираем все переменные (**Variable:** `All`) и нажимаем ОК для выполнения анализа.

3. Находим функцию **Distance Matrix** и открываем матрицу различий. При помощи главного меню **File > Save As...** сохраняем матрицу различий, присвоив ей имя. Обратите внимание: эта матрица совсем не похожа на нашу матрицу различий: мы ее используем только как готовый матричный формат.

4. Открываем файл, сохраненный на предыдущем шаге: он содержит необходимый нам матричный формат. Теперь необходимо заменить содержимое

этой матрицы на матрицу различий. Для этого опять копируем матрицу различий из программы Excel и переносим ее путем вставки в подготовленную матрицу. Сохраняем результат при помощи команды **Save**. Теперь можно приступать к кластерному анализу матрицы различий.

5. Открываем диалог метода кластерного анализа: **Statistics... > Cluster Analysis**. Выбираем **Joining (Tree clustering)**. Нажимаем ОК. Убеждаемся, что данные воспринимаются программой как матрица: **Input file: Distance matrix**.

6. Выбираем метод кластеризации. В поле **Amalgamation (linkage) rule** (Правило объединения) выбираем необходимый нам метод. Нажимаем ОК и получаем меню результатов **Joining Results**.

7. Нажимая кнопки с разными разделами результатов, просматриваем дендрограмму и таблицу последовательности агломерации. При необходимости нажимаем **Cancel** и возобновляем анализ с установкой другого метода.

КЛАСТЕРНЫЙ АНАЛИЗ И МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ

Многомерное шкалирование (глава 18) и кластерный анализ — это методы, основанные на *дистантной модели*: непосредственными исходными данными для них является информация о различии объектов. Поэтому представляет интерес сравнение результатов применения этих методов в отношении одних и тех же данных.

Напомним, что многомерное шкалирование, как и факторный анализ, направлено на выявление небольшого числа шкал. Эти шкалы трактуются как критерии, лежащие в основе различий объектов, и интерпретируются через объекты, поляризованные по этим шкалам. Особое значение при интерпретации шкал, в отличие от факторов, уделяется визуализации координатных представлений объектов в пространстве шкал, что связано с невозможностью поворота шкал относительно объектов (как факторов относительно признаков в факторном анализе). Поэтому особую ценность при шкалировании имеют двух-, максимум — трехшкальные решения. Так же, как и в факторном анализе, при многомерном шкалировании получение решения с малым числом шкал неизбежно влечет потерю исходной информации о различии объектов.

ПРИМЕР 19.3

Сравним результаты многомерного шкалирования и кластерного анализа одной и той же матрицы различий (табл. 19.3), полученной путем оцифровки исходной социометрической матрицы (табл. 19.2).

Предположим, исследователь решил применить многомерное шкалирование для того, чтобы по его результатам разместить 12 членов группы в аудитории в соответствии с их симпатиями и антипатиями. Графическое изображение 2-шкального решения приведено на рис. 19.7. Для сравнения на рис. 19.8 изображена дендрограмма — результат кластерного анализа методом средней связи для тех же данных.

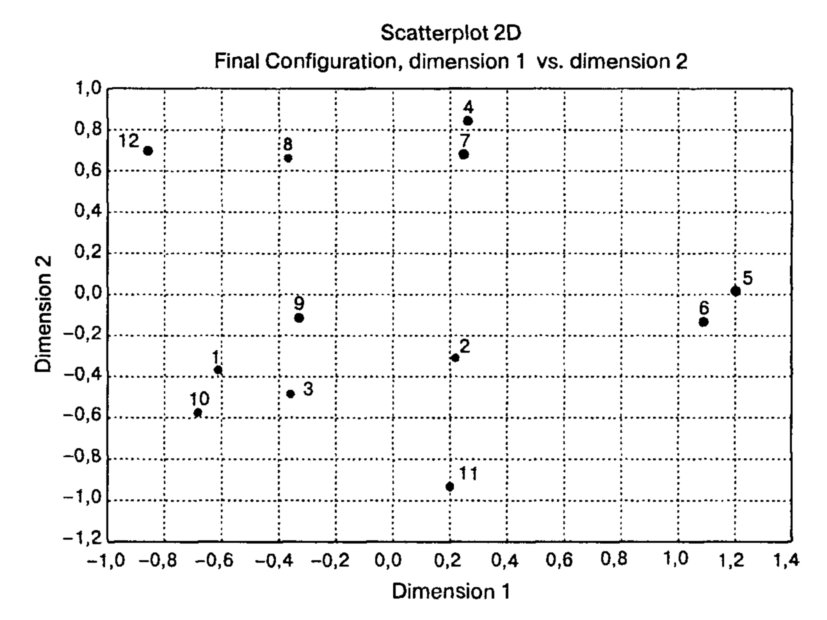


Рис. 19.7. Результат многомерного шкалирования данных социометрии (2-шкальное решение)

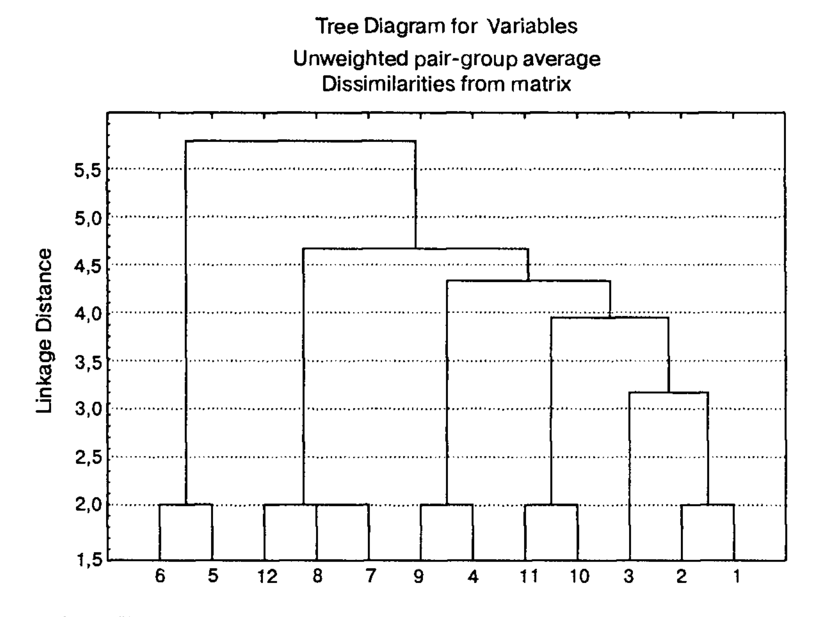


Рис.19.8. Дендрограмма для данных социометрии (метод средней связи)

Двумерная конфигурация объектов оказывается весьма грубым отражением реальных группировок. Так, группы (1, 2, 3) и (10, 11) на плоском изображении практически не разделены, а члены пары (9, 4) оказались в разных областях пространства. Несколько лучше оказывается 3-шкальное решение (рис. 19.9), хотя и оно весьма отдаленно напоминает группировки, выделяемые при помощи кластерного анализа.

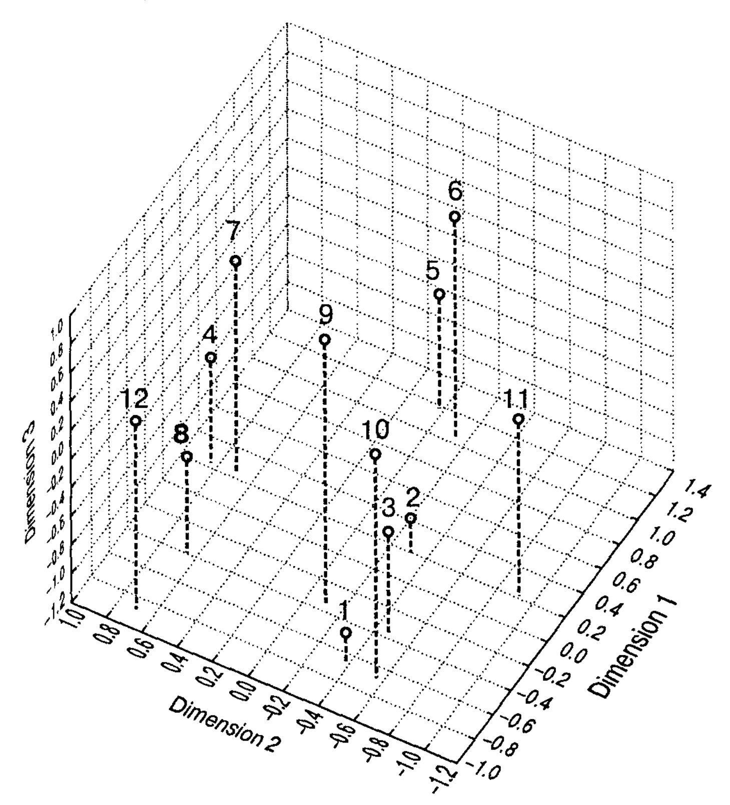


Рис. 19.9. Результат многомерного шкалирования данных социометрии (3-шкальное решение)

По-видимому, наблюдаемые искажения являются следствием того, что симпатии и антипатии в данной группе вряд ли могут быть объяснены небольшим числом общих оснований. Такое единодушие проявляется разве что в отношении очень небольшого числа членов этой группы (5 и 6).

Таким образом, наиболее серьезным ограничением многомерного шкалирования является, по-видимому, предположение о том, что в основе *всех* различий между объектами лежит небольшое число критериев (шкал). Если есть серьезные сомнения на этот счет, то применение многомерного шкалирования вряд ли оправданно. И тогда целесообразнее применять кластерный анализ.

Приложения

ОСНОВНЫЕ СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Приложение 1

СТАНДАРТНЫЕ НОРМАЛЬНЫЕ ВЕРОЯТНОСТИ

В таблице указаны значения площади под кривой единичного нормального распределения, находящиеся справа от *Z*. В крайнем левом столбце даны различные z-значения с точностью до одного десятичного знака. Значения вероятностей указаны для различных значений *Z*, включая второй знак после запятой (указан в верхнем ряду).

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4404	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007									
3,3	0,0005									
3,4	0,0003									
3,5	0,00023									
3,6	0,00016									
3,7	0,00011									
3,8	0,00007									
3,9	0,00005									
4,0	0,00003									

Значения рассчитаны при помощи программы Excel.

Приложение 2

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ *t*-СТЬЮДЕНТА (для проверки ненаправленных альтернатив — двусторонний критерий)

df	*p*				*df*	*p*			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
1	6,314	12,70	63,65	636,61	46	1,679	2,013	2,687	3,515
2	2,920	4,303	9,925	31,602	47	1,678	2,012	2,685	3,510
3	2,353	3,182	5,841	12,923	48	1,677	2,011	2,682	3,505
4	2,132	2,776	4,604	8,610	49	1,677	2,010	2,680	3,500
5	2,015	2,571	4,032	6,869	50	1,676	2,009	2,678	3,496
6	1,943	2,447	3,707	5,959	51	1,675	2,008	2,676	3,492
7	1,895	2,365	3,499	5,408	52	1,675	2,007	2,674	3,488
8	1,860	2,306	3,355	5,041	53	1,674	2,006	2,672	3,484
9	1,833	2,262	3,250	4,781	54	1,674	2,005	2,670	3,480
10	1,812	2,228	3,169	4,587	55	1,673	2,004	2,668	3,476
11	1,796	2,201	3,106	4,437	56	1,673	2,003	2,667	3,473
12	1,782	2,179	3,055	4,318	57	1,672	2,002	2,665	3,470
13	1,771	2,160	3,012	4,221	58	1,672	2,002	2,663	3,466
14	1,761	2,145	2,977	4,140	59	1,671	2,001	2,662	3,463
15	1,753	2,131	2,947	4,073	60	1,671	2,000	2,660	3,460
16	1,746	2,120	2,921	4,015	61	1,670	2,000	2,659	3,457
17	1,740	2,110	2,898	3,965	62	1,670	1,999	2,657	3,454
18	1,734	2,101	2,878	3,922	63	1,669	1,998	2,656	3,452
19	1,729	2,093	2,861	3,883	64	1,669	1,998	2,655	3,449
20	1,725	2,086	2,845	3,850	65	1,669	1,997	2,654	3,447
21	1,721	2,080	2,831	3,819	66	1,668	1,997	2,652	3,444
22	1,717	2,074	2,819	3,792	67	1,668	1,996	2,651	3,442
23	1,714	2,069	2,807	3,768	68	1,668	1,995	2,650	3,439
24	1,711	2,064	2,797	3,745	69	1,667	1,995	2,649	3,437
25	1,708	2,060	2,787	3,725	70	1,667	1,994	2,648	3,435
26	1,706	2,056	2,779	3,707	71	1,667	1,994	2,647	3,433
27	1,703	2,052	2,771	3,690	72	1,666	1,993	2,646	3,431
28	1,701	2,049	2,763	3,674	73	1,666	1,993	2,645	3,429

df	*p*				*df*	*p*			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
29	1,699	2,045	2,756	3,659	74	1,666	1,993	2,644	3,427
30	1,697	2,042	2,750	3,646	75	1,665	1,992	2,643	3,425
31	1,696	2,040	2,744	3,633	76	1,665	1,992	2,642	3,423
32	1,694	2,037	2,738	3,622	78	1,665	1,991	2,640	3,420
33	1,692	2,035	2,733	3,611	79	1,664	1,990	2,639	3,418
34	1,691	2,032	2,728	3,601	80	1,664	1,990	2,639	3,416
35	1,690	2,030	2,724	3,591	90	1,662	1,987	2,632	3,402
36	1,688	2,028	2,719	3,582	100	1,660	1,984	2,626	3,390
37	1,687	2,026	2,715	3,574	110	1,659	1,982	2,621	3,381
38	1,686	2,024	2,712	3,566	120	1,658	1,980	2,617	3,373
39	1,685	2,023	2,708	3,558	130	1,657	1,978	2,614	3,367
40	1,684	2,021	2,704	3,551	140	1,656	1,977	2,611	3,361
41	1,683	2,020	2,701	3,544	150	1,655	1,976	2,609	3,357
42	1,682	2,018	2,698	3,538	200	1,653	1,972	2,601	3,340
43	1,681	2,017	2,695	3,532	250	1,651	1,969	2,596	3,330
44	1,680	2,015	2,692	3,526	300	1,650	1,968	2,592	3,323
45	1,679	2,014	2,690	3,520	350	1,649	1,967	2,590	3,319

Значения рассчитаны при помощи программы Excel.

Приложение 3

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ *F*-ФИШЕРА, ДЛЯ ПРОВЕРКИ НАПРАВЛЕННЫХ АЛЬТЕРНАТИВ

$P = 0,05$

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	?
Степени свободы для знаменателя	3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,785	8,745	8,638	8,527
	5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,735	4,678	4,527	4,366
	7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,637	3,575	3,410	3,231
	10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	2,978	2,913	2,737	2,539
	11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,854	2,788	2,609	2,406
	12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,753	2,687	2,505	2,297
	13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,671	2,604	2,420	2,208
	14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,602	2,534	2,349	2,132
	15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,544	2,475	2,288	2,067
	16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,494	2,425	2,235	2,011
	18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,412	2,342	2,150	1,918
	20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,348	2,278	2,082	1,844
	30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,165	2,092	1,887	1,624
	40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,077	2,003	1,793	1,511
	50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,026	1,952	1,737	1,440
	70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	1,969	1,893	1,674	1,355
	100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,927	1,850	1,627	1,286
	200	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,878	1,801	1,572	1,192
	∞	3,843	2,998	2,607	2,374	2,216	2,100	2,011	1,940	1,833	1,754	1,519	

$P = 0,01$

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	?
Степени свободы для знаменателя	3	34,116	30,816	29,457	28,710	28,237	27,911	27,671	27,489	27,228	27,052	26,597	26,126
	5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,051	9,888	9,466	9,022
	7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,620	6,469	6,074	5,651
	10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,849	4,706	4,327	3,910
	11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,539	4,397	4,021	3,604
	12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,296	4,155	3,780	3,362
	13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,100	3,960	3,587	3,166
	14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	3,939	3,800	3,427	3,005
	15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,805	3,666	3,294	2,870
	16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,691	3,553	3,181	2,754
	18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,508	3,371	2,999	2,567
	20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,368	3,231	2,859	2,422
	30	7,562	5,390	4,510	4,018	3,699	3,473	3,305	3,173	2,979	2,843	2,469	2,008
	40	7,314	5,178	4,313	3,828	3,514	3,291	3,124	2,993	2,801	2,665	2,288	1,806
	50	7,171	5,057	4,199	3,720	3,408	3,186	3,020	2,890	2,698	2,563	2,183	1,685
	70	7,011	4,922	4,074	3,600	3,291	3,071	2,906	2,777	2,585	2,450	2,067	1,542
	100	6,895	4,824	3,984	3,513	3,206	2,988	2,823	2,694	2,503	2,368	1,983	1,429
	200	6,763	4,713	3,881	3,414	3,110	2,893	2,730	2,601	2,411	2,275	1,886	1,281
	∞	6,637	4,607	3,784	3,321	3,019	2,804	2,641	2,513	2,323	2,187	1,793	

Значения рассчитаны при помощи программы Excel.

Приложение 4

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ χ^2

df	*p*				*df*	*p*			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
1	2,706	3,842	6,635	10,829	46	58,641	62,841	71,221	81,431
2	4,605	5,992	9,211	13,817	47	59,774	64,013	72,463	82,752
3	6,251	7,815	11,346	16,269	48	60,907	65,183	73,703	84,069
4	7,779	9,488	13,278	18,470	49	62,038	66,351	74,940	85,384
5	9,236	11,071	15,088	20,519	50	63,167	67,518	76,175	86,694
6	10,645	12,593	16,814	22,462	51	64,295	68,683	77,408	88,003
7	12,017	14,068	18,478	24,327	52	65,422	69,846	78,638	89,308
8	13,362	15,509	20,093	26,130	53	66,548	71,008	79,866	90,609
9	14,684	16,921	21,669	27,883	54	67,673	72,168	81,092	91,909
10	15,987	18,309	23,213	29,594	55	68,796	73,326	82,316	93,205
11	17,275	19,677	24,729	31,271	56	69,919	74,484	83,538	94,499
12	18,549	21,028	26,221	32,917	57	71,040	75,639	84,758	95,790
13	19,812	22,365	27,693	34,536	58	72,160	76,794	85,976	97,078
14	21,064	23,688	29,146	36,132	59	73,279	77,947	87,192	98,365
15	22,307	24,999	30,583	37,706	60	74,397	79,099	88,406	99,649
16	23,542	26,299	32,006	39,262	61	75,514	80,232	89,591	100,887
17	24,769	27,591	33,415	40,801	62	76,630	81,381	90,802	102,165
18	25,989	28,873	34,812	42,323	63	77,745	82,529	92,010	103,442
19	27,204	30,147	36,198	43,832	64	78,860	83,675	93,217	104,717
20	28,412	31,415	37,574	45,327	65	79,973	84,821	94,422	105,988
21	29,615	32,675	38,940	46,810	66	81,085	85,965	95,626	107,257
22	30,813	33,929	40,298	48,281	67	82,197	87,108	96,828	108,525
23	32,007	35,177	41,647	49,742	68	83,308	88,250	98,028	109,793
24	33,196	36,420	42,989	51,194	69	84,418	89,391	99,227	111,055
25	34,382	37,658	44,324	52,635	70	85,527	90,531	100,425	112,317
26	35,563	38,891	45,652	54,068	71	86,635	91,670	101,621	113,577
27	36,741	40,119	46,973	55,493	72	87,743	92,808	102,816	114,834
28	37,916	41,343	48,289	56,910	73	88,850	93,945	104,010	116,092
29	39,087	42,564	49,599	58,320	74	89,956	95,081	105,202	117,347
30	40,256	43,780	50,904	59,722	75	91,061	96,217	106,393	118,599
31	41,422	44,993	52,203	61,118	76	92,166	97,351	107,582	119,850
32	42,585	46,202	53,498	62,508	78	94,374	99,617	109,958	122,347

df	*p*				*df*	*p*			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
33	43,745	47,408	54,789	63,891	79	95,476	100,749	111,144	123,595
34	44,903	48,610	56,074	65,269	80	96,578	101,879	112,329	124,839
35	46,059	49,810	57,356	66,641	90	107,565	113,145	124,116	137,208
36	47,212	51,007	58,634	68,008	100	118,498	124,342	135,807	149,449
37	48,363	52,201	59,907	69,370	110	129,385	135,480	147,414	161,582
38	49,513	53,393	61,177	70,728	120	140,233	146,567	158,950	173,618
39	50,660	54,582	62,444	72,080	130	151,045	157,610	170,423	185,573
40	51,805	55,768	63,707	73,428	140	161,827	168,613	181,841	197,450
41	52,949	56,953	64,967	74,772	150	172,581	179,581	193,207	209,265
42	54,090	58,135	66,224	76,111	200	226,021	233,994	249,445	267,539
43	55,230	59,314	67,477	77,447	250	279,050	287,882	304,939	324,831
44	56,369	60,492	68,728	78,779	300	331,788	341,395	359,906	381,424
45	57,505	61,668	69,976	80,107	350	384,306	394,626	414,474	437,487

Значения рассчитаны при помощи программы Excel.

Приложение 5

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ ДЛЯ ЧИСЛА СЕРИЙ
(для проверки ненаправленных гипотез для $\alpha = 0,05$)

$W_{0,025}$

N_2 \ N_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4																			
5			2	2															
6		2	2	3	3														
7		2	2	3	3	3													
8		2	3	3	3	4	4												
9		2	3	3	4	4	5	5											
10		2	3	3	4	5	5	5	6										
11		2	3	4	4	5	5	6	6	7									
12	2	2	3	4	4	5	6	6	7	7	7								
13	2	2	3	4	5	5	6	6	7	7	8	8							
14	2	2	3	4	5	5	6	7	7	8	8	9	9						
15	2	3	3	4	5	6	6	7	7	8	8	9	9	10					
16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11				
17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11			
18	2	3	4	5	5	6.	7	8	8	9	9	10	10	11	11	12	12		
19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	
20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

$W_{0,975}$

N_2 \ N_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1																			
2	4																		
3	5	6																	
4	5	7	8																
5	5	7	8	9															
6	5	7	8	9	10														
7	5	7	9	10	11	12													
8	5	7	9	10	11	12	13												
9	5	7	9	11	12	13	13	14											
10	5	7	9	11	12	13	14	15	15										
11	5	7	9	11	12	13	14	15	16	16									
12	5	7	9	11	12	13	15	15	16	17	18								
13	5	7	9	11	13	14	15	16	17	18	18	19							
14	5	7	9	11	13	14	15	16	17	18	19	19	20						
15	5	7	9	11	13	14	15	17	17	18	19	20	21	21					
16	5	7	9	11	13	15	16	17	18	19	20	20	21	22	23				
17	5	7	9	11	13	15	16	17	18	19	20	21	22	22	23	24			
18	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	24	25		
19	5	7	9	11	13	15	16	17	19	20	21	22	22	23	24	25	25	26	
20	5	7	9	11	13	15	16	17	19	20	21	22	23	24	24	25	26	26	27

$N_1 = N_2$	$W_{0,025}$	$W_{0,375}$	$N_1 = N_2$	$W_{0,025}$	$W_{0,975}$
20	14	27	40	31	50
21	15	28	42	33	52
22	16	29	44	35	54
23	16	31	46	37	56
24	17	32	48	38	59
25	18	33	50	40	61
26	19	34	55	45	66
27	20	35	60	49	72
2S.	21	36	65	54	77
29	22	37	70	58	83
30	22	39	75	63	88
32	24	41	80	68	93
34	26	43	85	72	99
36	28	45	90	77	104
38	30	47	95	82	109
			100	86	115

Примечание: гипотеза H_0 отклоняется, если эмпирическое число серий W меньше или равно $W_{0,025}$ (больше или равно $W_{0,975}$).

Источник: Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. М., 1980. С. 547–548.

Приложение 6

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ *r*-ПИРСОНА (*r*-СПИРМЕНА)

(для проверки ненаправленных альтернатив, *n* — объем выборки)

n	*p*				*n*	*p*			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428
16	0,426	0,497	0,623	0,742	57	0,220	0,261	0,339	0,424
17	0,412	0,482	0,606	0,725	58	0,218	0,259	0,336	0,421
18	0,400	0,468	0,590	0,708	59	0,216	0,256	0,333	0,418
19	0,389	0,456	0,575	0,693	60	0,214	0,254	0,330	0,414
20	0,378	0,444	0,561	0,679	61	0,213	0,252	0,327	0,411
21	0,369	0,433	0,549	0,665	62	0,211	0,250	0,325	0,408
22	0,360	0,423	0,537	0,652	63	0,209	0,248	0,322	0,405
23	0,352	0,413	0,526	0,640	64	0,207	0,246	0,320	0,402
24	0,344	0,404	0,515	0,629	65	0,206	0,244	0,317	0,399
25	0,337	0,396	0,505	0,618	66	0,204	0,242	0,315	0,396
26	0,330	0,388	0,496	0,607	67	0,203	0,240	0,313	0,393
27	0,323	0,381	0,487	0,597	68	0,201	0,239	0,310	0,390
28	0,317	0,374	0,479	0,588	69	0,200	0,237	0,308	0,388
29	0,311	0,367	0,471	0,579	70	0,198	0,235	0,306	0,385
30	0,306	0,361	0,463	0,570	80	0,185	0,220	0,286	0,361
31	0,301	0,355	0,456	0,562	90	0,174	0,207	0,270	0,341

n	p				n	p			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
32	0,296	0,349	0,449	0,554	100	0,165	0,197	0,256	0,324
33	0,291	0,344	0,442	0,547	110	0,158	0,187	0,245	0,310
34	0,287	0,339	0,436	0,539	120	0,151	0,179	0,234	0,297
35	0,283	0,334	0,430	0,532	130	0,145	0,172	0,225	0,285
36	0,279	0,329	0,424	0,525	140	0,140	0,166	0,217	0,275
37	0,275	0,325	0,418	0,519	150	0,135	0,160	0,210	0,266
38	0,271	0,320	0,413	0,513	200	0,117	0,139	0,182	0,231
39	0,267	0,316	0,408	0,507	250	0,104	0,124	0,163	0,207
40	0,264	0,312	0,403	0,501	300	0,095	0,113	0,149	0,189
41	0,260	0,308	0,398	0,495	350	0,088	0,105	0,138	0,175
42	0,257	0,304	0,393	0,490	400	0,082	0,098	0,129	0,164
43	0,254	0,301	0,389	0,484	450	0,078	0,092	0,121	0,155
44	0,251	0,297	0,384	0,479	500	0,074	0,088	0,115	0,147
45	0,248	0,294	0,380	0,474	600	0,067	0,080	0,105	0,134

Значения рассчитаны при помощи программы Excel.

Приложение 7

ЗНАЧЕНИЯ *Z*-ПРЕОБРАЗОВАНИЯ ФИШЕРА ДЛЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ

r_{ij}	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
2	0, 2027	0,2132	0, 2237	0,2342	0,2448	0,2554	0, 2661	0, 2769	0, 2877	0, 2986
3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
5	0, 5493	0,5627	0, 5763	0, 5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
6	0,6931	0,7089	0,7250	0, 7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Значения рассчитаны при помощи программы Excel.

Приложение 8

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ *F*-ФИШЕРА ДЛЯ ПРОВЕРКИ НЕНАПРАВЛЕННЫХ АЛЬТЕРНАТИВ

$P = 0,05$

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	?
Степени свободы для знаменателя	3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,419	14,337	14,124	13,903
	5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,619	6,525	6,278	6,017
	7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,761	4,666	4,415	4,144
	10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,717	3,621	3,365	3,081
	11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,526	3,430	3,173	2,884
	12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,374	3,277	3,019	2,726
	13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,250	3,153	2,893	2,597
	14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,147	3,050	2,789	2,489
	15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,060	2,963	2,701	2,397
	16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	2,986	2,889	2,625	2,318
	18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,866	2,769	2,503	2,189
	20	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,774	2,676	2,408	2,087
	30	5,568	4,182	3,589	3,250	3,026	2,867	2,746	2,651	2,511	2,412	2,136	1,789
	40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,388	2,288	2,007	1,639
	50	5,340	3,975	3,390	3,054	2,833	2,674	2,553	2,458	2,317	2,216	1,931	1,548
	70	5,247	3,890	3,309	2,975	2,754	2,595	2,474	2,379	2,237	2,136	1,847	1,438
	100	5,179	3,828	3,250	2,917	2,696	2,537	2,417	2,321	2,179	2,077	1,784	1,351
	200	5,100	3,758	3,182	2,850	2,630	2,472	2,351	2,256	2,113	2,010	1,712	1,233
	?	5,027	3,692	3,119	2,788	2,569	2,411	2,290	2,194	2,051	1,947	1,643	

$P = 0{,}01$

Степени свободы для знаменателя	Степени свободы для числителя											
	1	2	3	4	5	6	7	8	10	12	24	?
3	55,552	49,800	47,468	46,195	45,391	44,838	44,434	44,125	43,685	43,387	42,623	41,833
5	22,785	18,314	16,530	15,556	14,939	14,513	14,200	13,961	13,618	13,385	12,780	12,147
7	16,235	12,404	10,883	10,050	9,522	9,155	8,885	8,678	8,380	8,176	7,645	7,079
10	12,827	9,427	8,081	7,343	6,872	6,545	6,303	6,116	5,847	5,661	5,173	4,641
11	12,226	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,418	5,236	4,756	4,228
12	11,754	8,510	7,226	6,521	6,071	5,757	5,524	5,345	5,085	4,906	4,431	3,907
13	11,374	8,186	6,926	6,233	5,791	5,482	5,253	5,076	4,820	4,643	4,173	3,649
14	11,060	7,922	6,680	5,998	5,562	5,257	5,031	4,857	4,603	4,428	3,961	3,439
15	10,798	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,424	4,250	3,786	3,263
16	10,576	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,272	4,099	3,638	3,114
18	10,218	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,030	3,860	3,402	2,876
20	9,944	6,987	5,818	5,174	4,762	4,472	4,257	4,090	3,847	3,678	3,222	2,693
30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,344	3,179	2,727	2,179
40	8,828	6,066	4,976	4,374	3,986	3,713	3,509	3,350	3,117	2,953	2,502	1,935
50	8,626	5,902	4,826	4,232	3,849	3,579	3,376	3,219	2,988	2,825	2,373	1,790
70	8,403	5,720	4,661	4,076	3,698	3,431	3,232	3,076	2,846	2,684	2,231	1,622
100	8,241	5,589	4,542	3,963	3,589	3,325	3,127	2,972	2,744	2,583	2,128	1,490
200	8,057	5,441	4,408	3,837	3,467	3,206	3,010	2,856	2,629	2,468	2,012	1,320
?	7,886	5,304	4,284	3,720	3,355	3,096	2,901	2,749	2,523	2,363	1,903	

Значения рассчитаны при помощи программы Excel.

Приложение 9

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ *U*-МАННА-УИТНИ

(для проверки ненаправленных альтернатив)

$P = 0{,}05$

N_2	N_1													
	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	34	41	48	55	62	69	76	83	90	98	105	112	119	127

$P = 0{,}01$

N_2	N_1													
	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3			0	0	0	1	1	1	2	2	2	2	3	3
4	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	1	2	3	4	4	6	7	7	8	9	10	11	12	13
6	3	4	5	6	6	9	10	11	12	13	15	16	17	18
7	4	6	7	9	9	12	13	15	16	18	19	21	22	24
8	6	7	9	11	11	15	17	18	20	22	24	26	28	30
9	7	9	11	13	13	18	20	22	24	27	29	31	33	36
10	9	11	13	16	16	21	24	26	29	31	34	37	39	42
11	10	13	16	18	18	24	27	30	33	36	39	42	45	48
12	12	15	18	21	21	27	31	34	37	41	44	47	51	54
13	13	17	20	24	24	31	34	38	42	45	49	53	56	60
14	15	18	22	26	26	34	38	42	46	50	54	58	63	67
15	16	20	24	29	29	37	42	46	51	55	60	64	69	73
16	18	22	27	31	31	41	45	50	55	60	65	70	74	79
17	19	24	29	34	34	44	49	54	60	65	70	75	81	86
18	21	26	31	37	37	47	53	58	64	70	75	81	87	92
19	22	28	33	39	39	51	56	63	69	74	81	87	93	99
20	24	30	36	42	42	54	60	67	73	79	86	92	99	105

Источник: Мартин Д. Психологические эксперименты. СПб, 2002. С. 453–454

Приложение 10

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ *T*-ВИЛКОКСОНА
(для проверки ненаправленных альтернатив)

n	Уровень значимости для одностороннего критерия				*n*	Уровень значимости для одностороннего критерия			
	0,05	0,025	0,01	0,005		0,05	0,025	0,01	0,005
	Уровень значимости для двустороннего критерия					Уровень значимости для двустороннего критерия			
	0,10	0,05	0,02	0,01		0,10	0,05	0,02	0,01
5	0				28	130	116	101	91
6	2	0	—	—	29	140	126	110	100
7	3	2	0	—	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

Источник: Мартин Д. Психологические эксперименты. СПб, 2002. С. 457

Приложение 11

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ *G* ЗНАКОВ

(для проверки ненаправленных альтернатив)

n	*p*		*n*	*p*		*n*	*p*		*n*	*p*	
	0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01
5	0	—	27	8	7	49	18	15	92	37	34
6	0	—	28	8	7	50	18	16	94	38	35
7	0	0	29	9	7	52	19	17	96	39	36
8	1	0	30	10	8	54	20	18	98	40	37
9	1	0	31	10	8	56	21	18	100	41	37
10	1	0	32	10	8	58	22	19	110	45	42
11	2	1	33	11	9	60	23	20	120	50	46
12	2	1	34	11	9	62	24	21	130	55	51
13	3	1	35	12	10	64	24	22	140	59	55
14	3	2	36	12	10	66	25	23	150	64	60
15	3	2	37	13	10	68	26	23	160	69	64
16	4	2	38	13	11	70	27	24	170	73	69
17	4	3	39	13	11	72	28	25	180	78	73
18	5	3	40	14	12	74	29	26	190	83	78
19	5	4	41	14	12	76	30	27	200	87	83
20	5	4	42	15	13	78	31	28	220	97	92
21	6	4	43	15	13	80	32	29	240	106	101
22	6	5	44	16	13	82	33	30	260	116	110
23	7	5	45	16	14	84	33	30	280	125	120
24	7	5	46	16	14	86	34	31	300	135	129
25	7	6	47	17	15	88	35	32			
26	8	6	48	17	15	90	36	33			

Источник: Сидоренко Е. Методы математической обработки в психологии. СПб, 2002. С. 323

Приложение 12

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ H-КРАСКАЛА-УОЛЛЕСА ДЛЯ ТРЕХ ВЫБОРОК ЧИСЛЕННОСТЬЮ $n \leqslant 5$

(для проверки ненаправленных альтернатив)

Размеры выборки			H	P	Размеры выборки			H	P
n_1	n_2	n_3			n_1	n_2	n_3		
2	1	1	2,7000	0,500	4	3	2	6,4444	0,008
								6,3000	0,011
2	2	1	3,6000	0,200				5,4444	0,046
								5,4000	0,051
2	2	2	4,5714	0,067				4,5111	0,098
			3,7143	0,200				4,4444	0,102
3	1	1	3,2000	0,300	4	3	3	6,7455	0,010
								6,7091	0,013
3	2	1	4,2857	0,100				5,7909	0,046
			3,8571	0,133				5,7273	0,050
								4,7091	0,092
3	2	2	5,3572	0,029				4,7000	0,101
			4,7143	0,048					
			4,5000	0,067	4	4	1	6,6667	0,010
			4,4643	0,105				6,1667	0,022
								4,9667	0,048
3	3	1	5,1429	0,043				4,8667	0,054
			4,5714	0,100				4,1667	0,082
			4,0000	0,129				4,0667	0,102
3	3	2	6,2500	0,011	4	4	2	7,0364	0,006
			5,3611	0,032				6,8727	0,011
			5,1389	0,061				5,4545	0,046
			4,5556	0,100				5,2364	0,052
			4,2500	0,121				4,5545	0,098
								4,4455	0,103
3	3	3	7,2000	0,004					

Размеры выборки			H	P
n_1	n_2	n_3		
			6,4889	0,011
			5,6889	0,029
			5,6000	0,050
			5,0667	0,086
			4,6222	0,100
4	1	1	3,5714	0,200
4	2	1	4,8214	0,057
			4,5000	0,076
			4,0179	0,114
4	2	2	6,0000	0,014
			5,3333	0,033
			5,1250	0,052
			4,4583	0,100
			4,1667	0,105
4	3	1	5,8333	0,021
			5,2083	0,050
			5,0000	0,057
			4,0556	0,093
			3,8889	0,129
5	2	2	6,5333	0,008
			6,1333	0,013
			5,1600	0,034
			5,0400	0,056
			4,3733	0,090
			4,2933	0,122
5	3	1	6,4000	0,012
			4,9600	0,048
			4,8711	0,052
			4,0178	0,095
			3,8400	0,123
5	3	2	6,9091	0,009
			6,8218	0,010
			5,2509	0,049
			5,1055	0,052
			4,6509	0,091
			4,4945	0,101
4	4	3	7,1439	0,010
			7,1364	0,011
			5,5985	0,049
			5,5758	0,051
			4,5455	0,099
			4,4773	0,102
4	4	4	7,6538	0,008
			7,5385	0,011
			5,6923	0,049
			5,6538	0,054
			4,6539	0,097
			4,5001	0,104
5	1	1	3,8571	0,143
5	2	1	5,2500	0,036
			5,0000	0,048
			4,4500	0,071
			4,2000	0,095
			4,0500	0,119
5	4	4	7,7604	0,009
			7,7440	0,011
			5,6571	0,049
			5,6176	0,050
			4,6187	0,100
			4,5527	0,102
5	5	1	7,3091	0,009
			6,8364	0,011
			5,1-273	0,046
			4,9091	0,053
			4,1091	0,086
			4,0364	0,105
5	5	2	7,3385	0,010
			7,2692	0,010
			5,3385	0,047
			5,2462	0,051
			4,6231	0,097
			4,5077	0,100

Размеры выборки			*H*	*P*	Размеры выборки			*H*	*P*
n_1	n_2	n_3			n_1	n_2	n_3		
5	3	3	7,0788	0,009	5	5	3	7,5780	0,010
			6,9818	0,011				7,5429	0,010
			5,6485	0,049				5,7055	0,046
			5,5152	0,051				5,6264	0,051
			4,5333	0,097				4,5451	0,100
			4,4121	0,109					
5	4	1	6,9545	0,008	5	5	4	7,8229	0,010
			6,8400	0,011				7,7914	0,010
			4,9855	0,044				5,6657	0,049
			4,8600	0,056				5,6429	0,050
			3,9873	0,098				4,5229	0,099
			3,9600	0,102				4,5200	0,101
5	4	2	7,2045	0,009	5	5	5	8,0000	0,009
			7,1182	0,010				7,9800	0,010
			5,2727	0,049				5,7800	0,049
			5,2682	0,050				5,6600	0,051
			4,5409	0,098				4,5600	0,100
			4,5182	0,101				4,5000	0,102
5	4	3	7,4449	0,010					
			7,3949	0,011					
			5,6564	0,049					
			5,6308	0,050					
			4,5487	0,099					
			4,5231	0,103					

Источник: Мартин Д. Психологические эксперименты. СПб, 2002. С. 460–462

Приложение 13

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ χ^2-ФРИДМАНА ДЛЯ ТРЕХ ВЫБОРОК ЧИСЛЕННОСТЬЮ n < 10

(для проверки ненаправленных альтернатив)

$n = 2$		$n = 3$		$n = 4$		$n = 5$	
χ^2_r	p	χ^2_r	p	χ^2_r	p	χ^2_r	p
0	1,000	0,000	1,000	0,0	1,000	0,0	1,000
1	0,833	0,667	0,944	0,5	0,931	0,4	0,954
3	0,509	2,000	0,528	1,5	0,653	1,2	0,691
4	0,167	2,667	0,361	2,0	0,431	1,6	0,522
		4,667	0,194	3,5	0,273	2,8	0,367
		6,000	0,028	4,5	0,125	3,6	0,182
				6,0	0,069	4,8	0,124
				6,5	0,042	5,2	0,093
				8,0	0,0046	6,4	0,039
						7,6	0,024
						8,4	0,0085
						10,0	0,00077

$n = 6$		$n = 7$		$n = 8$		$n = 9$	
χ^2_r	p	χ^2_r	p	χ^2_r	p	χ^2_r	p
0,00	1,000	0,000	1,000	0,00	1,000	0,000	1,000
0,33	0,956	0,286	0,964	0,25	0,967	0,222	0,971
1,00	0,740	0,857	0,768	0,75	0,794	0,667	0,814
1,33	0,570	1,143	0,620	1,00	0,654	0,889	0,865
2,33	0,430	2,000	0,486	1,75	0,531	1,556	0,569
3,00	0,252	2,571	0,305	2,25	0,355	2,000	0,398
4,00	0,184	3,429	0,237	3,00	0,285	2,667	0,328
4,33	0,142	3,714	0,192	3,25	0,236	2,889	0,278
5,33	0,072	4,571	0,112	4,00	0,149	3,556	0,187
6,33	0,052	5,429	0,085	4,75	0,120	4,222	0,154
7,00	0,029	6,000	0,052	5,25	0,079	4,667	0,107
8,33	0,012	7,143	0,027	6,25	0,047	5,556	0,069
9,00	0,0081	7,714	0,021	6,75	0,038	6,000	0,057

$n = 6$		$n = 7$		$n = 8$		$n = 9$	
χ^2_r	p	χ^2_r	p	χ^2_r	p	χ^2_r	p
9,33	0,0055	8,000	0,016	7,00	0,030	6,222	0,048
10,33	0,0017	8,857	0,0084	7,75	0,018	6,889	0,031
12,00	0,00013	10,286	0,0036	9,00	0,0099	8,000	0,019
		10,571	0,0027	9,25	0,0080	8,222	0,016
		11,143	0,0012	9,75	0,0048	8,667	0,010
		12,286	0,00032	10,75	0,0024	9,556	0,0060
		14,000	0,000021	12,00	0,0011	10,667	0,0035
				12,25	0,00086	10,889	0,0029
				13,00	0,00026	11,556	0,0013
				14,25	0,000061	12,667	0,00066
				16,00	0,0000036	13,556	0,00035
						14,000	0,00020
						14,222	0,000097
						14,889	0,000054
						16,222	0,000011
						18,000	0,0000006

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ χ^2-ФРИДМАНА ДЛЯ ЧЕТЫРЕХ ВЫБОРОК ЧИСЛЕННОСТЬЮ $n < 5$

$n = 2$		$n = 3$		$n = 4$			
χ^2_r	p	χ^2_r	p	χ^2_r	p	χ^2_r	p
0,0	1,000	0,0	1,000	0,0	1,000	5,7	0,141
0,6	0,958	0,6	0,958	0,3	0,992	6,0	0,105
1,2	0,834	1,0	0,910	0,6	0,928	6,3	0,094
1,8	0,792	1,8	0,727	0,9	0,900	6,6	0,077
2,4	0,625	2,2	0,608	1,2	0,800	6,9	0,068
3,0	0,542	2,6	0,524	1,5	0,754	7,2	0,054
3,6	0,458	3,4	0,446	1,8	0,677	7,5	0,052
4,2	0,375	3,8	0,342	2,1	0,649	7,8	0,036
4,8	0,208	4,2	0,300	2,4	0,524	8,1	0,033
5,4	0,167	5,0	0,207	2,7	0,508	8,4	0,019
6,0	0,042	5,4	0,175	3,0	0,432	8,7	0,014
		5,8	0,148	3,3	0,389	9,3	0,012
		6,6	0,075	3,6	0,355	9,6	0,0069
		7,0	0,054	3,9	0,324	9,9	0,0062
		7,4	0,033	4,5	0,242	10,2	0,0027
		8,2	0,017	4,8	0,200	10,8	0,0016
		9,0	0,0017	5,1	0,190	11,1	0,00094
				5,4	0,158	12,0	0,000072

Источник: Сидоренко Е. Методы математической обработки в психологии. СПб, 2002. С. 325–326

АНГЛО-РУССКИЙ ТЕРМИНОЛОГИЧЕСКИЙ СЛОВАРЬ

1-Sample K-S Test — критерий (тест) Колмогорова–Смирнова
1-tailed — односторонний (направленный) уровень значимости
2-tailed — двусторонний (ненаправленный) уровень значимости
Absolute value — абсолютное значение
Actual (value, group) — действительное, реальное (значение, группа)
Add — добавить
Adjusted — исправленный (улучшенный)
Advanced (Model) — специальная, более совершенная (модель)
Agglomeration schedule — последовательность агломерации (объединения)
ALSCAL — программа неметрического многомерного шкалирования
Amalgamation — слияние, объединение
Analyze — анализировать
ANOVA — дисперсионный анализ
Approach — подход
Assume — принятие (допущение, предположение)
Asymmetric — асимметричная
Asymp. Sig. — примерный (приближенный) уровень значимости
Average Linkage — средней связи (метод кластеризации)
Averaged — усредненный
Axis — ось (координат)
Bartlett's Test of Sphericity — тест сферичности Бартлета
Based on — основанный на (исходящий из)
Beta-Coefficient — стандартизированный коэффициент регрессии
Between (objects, variables) — между (объектами, переменными)
Between Groups Linkage — межгрупповой (средней) связи (метод кластеризации)
Between-Group — межгрупповой
Between-Subject — между объектами (межгрупповой)
Binary Measures — количественные показатели (меры) для бинарных данных
Binomial Test — биномиальный критерий
Bivariate — двумерный
Box's M–test — M–тест Бокса
Canonical Analysis — канонический анализ
Case — случай (испытуемый)
Casewise deletion — исключение из анализа случая (строки), в котором имеется пропуск хотя бы одного значения
Categories — категории (номинативного признака)
Categorization — операция выделения интервалов квантования (или значений переменной) при построении гистограммы и составлении таблицы частот
Cell — ячейка (таблицы)
Central Tendency — центральная тенденция (мера)
Centroid — центроид
Chi I — хи–квадрат
Chi-square (Test) — хи–квадрат (критерий)
Classify — классифицировать
Cluster Combined — объединенные кластеры
Cluster Method — метод кластеризации
Coefficient(s) — коэффициент(ы)
Column — столбец
Combine — объединение, объединять
Communality — общность
Compare — сравнивать
Compare Means — сравнение средних
Comparison — сравнение
Complete Linkage — полной связи (метод кластеризации)
Compute — вычисление, вычислять
Conditionality — условность, обусловленность (подгонки)

Confidence (Interval) — доверительный (интервал)
Constant — константа
Contrast — контраст
Controlling for... — контролировать (фиксировать) для...
Convergence — сходимость (при подгонки)
Corrected (Model) — исправленная, скорректированная (модель)
Correlate — коррелировать (определять совместную изменчивость)
Correlation matrix — корреляционная матрица
Count Measure — количественный показатель (мера) частоты
Covariance — ковариация
Covariate — ковариата
Criteria (Criterion) — условие (критерий)
Crosstabulation (Crosstab) — сопряженность, кросстабуляция
Cumulative frequencies — кумулятивные (накопленные) частоты
Custom Model — специальная модель
Cut Point — точка деления
Data — данные
Data Editor — редактор (таблица) исходных данных в SPSS
Data Reduction — сокращение данных (метод)
Define (Groups) — определение, задание (групп)
Degrees of freedom (df) — число степеней свободы
Deletion — удаление (исключение)
Dendrogram — дендрограмма (древовидный график)
Density Function — функция плотности вероятности
Dependent Sample — зависимая выборка
Dependent-Samples T Test — критерий t-Стьюдента для зависимых выборок
Derived — производный
Descriptive Statistics — описательные статистики
df — число степеней свободы (сокр.)
Difference — разность, различие
Dimension — шкала
Discriminant Analysis — дискриминантный анлиз
Dispersion — изменчивость
Dissimilarity — различие
Distance — расстояние
Distribution — распределение
Distribution Function — функция распределения (вероятности)
Effect — влияние (фактора)
Eigenvalue — собственное значение
Enter — исходный (метод)
Entry — включение
Epsilon Corrected — с Эпсилон-коррекцией
Equal — одинаковые
Equal Variances — одинаковые (эквивалентные) дисперсии
Equality (of Variances) — эквивалентность, равенство (дисперсии)
Error — ошибка
Estimate — оценка
Euclidean Distance — Евклидово расстояние
Exact — точный, точно
Exact Sig. — точный уровень значимости
Exclude (cases) — исключение (объектов)
Expected Frequency (Value) — ожидаемая (теоретическая) частота (значение)
Explained (Variance) — «объясненная» (дисперсия)
Extraction Method — метод факторизации (экстракции факторов)
Factor — фактор
Factor Analysis — факторный анализ
Factor Loadings — факторные нагрузки
Factor Score coefficients — коэффициенты факторных оценок
Factor Scores — факторные оценки
Fisher's Exact Test — точный критерий Фишера
Fixed (Factor) — постоянный, фиксированный (фактор)
For Matrix — для всей матрицы
F-ratio Variances — значение F-критерия (F-отношения) для дисперсий
Frequency — частота
Frequency tables — таблицы частот
Friedman test (ANOVA) — критерий Фридмана
Full factorial model — модель, включающая все факторы
Furthest Neighbor — дальнего соседа (метод кластеризации)
General Linear Model — общая линейная модель
Generalized least squares — обобщенный метод наименьших квадратов
GLM — общая линейная модель (сокр.)
Goodness-of-fit — качество подгонки
Group Membership — принадлежность к группе
Group Plots — график(и) для всей группы
Grouping Variable — группирующая переменная
Hierarchical Cluster Analysis — иерархический кластерный анализ
Histogram — гистограмма
Homogeneity of Variances — гомогенность (однородность) дисперсий
Horizontal — горизонтальная (ориентация)
Image Factoring — факторный анализ образов
Improvement — улучшение (с поправкой)

Independent Sample — независимая выборка
Independent-Samples T Test — критерий t-Стьюдента для независимых выборок
Individual Differences (model) — индивидуальных различий (модель)
INDSCALE — программа многомерного шкалирования индивидуальных различий
Initial (conditions) — начальные (условия)
Input File — исходный, входящий файл
Interaction — взаимодействие
Intercept — свободный член (уравнения)
Interval — интервальная (шкала)
Interval data — данные в интервальной шкале
Interval Scale — интервальная шкала
Inverse distribution function — обратная функция распределения
Iteration — итерация
Iteration history — «история» (последовательность) итераций
Iterations stopped... — итерации остановлены...
Joining — объединение, связь
Kendall's tau — тау-Кендалла (корреляция)
Kendall's tau-b — тау-b Кендалла (корреляция)
Kruskal-Wallis H — критерий H-Краскала Уоллеса
Kruskal-Wallis one-way analysis of variance — однофакторный дисперсионный анализ Краскала–Уоллеса
Kurtosis — эксцесс
Label — метка, обозначение
Level — уровень
Level of Measurement — уровень (шкала) измерения
Levene's Test — критерий Ливена
Linear Regression — линейный регрессионный
Linkage — соединение, связь
List — список
Listwise — построчно
Listwise Deletion — исключение из анализа случая (строки), в котором имеется пропуск хотя бы одного значения
Loading — нагрузка
Mann-Whitney U — критерий U-Манна-Уитни
MANOVA — многомерный дисперсионный анализ
Marginal (Means) — отдельные (средние значения)
Matrices — матрицы
Matrix — матрица
Mauchly's Test of Sphericity — тест сферичности Моучли
Maximum Likelihood –максимального правдоподобия (метод)
McNemar Test — критерий Мак-Нимара
Mean — среднее
Mean of Squares (MS) — средний квадрат
Mean Substitution — замена пропущенных значений средними
Means Plot — график средних значений
Measurement — измерение
Median — медиана
Missing (Values) — пропущенные (значения)
M-L (Maximum likelihood) — максимального правдоподобия (метод, оценка)
Mode — мода
Monte Carlo Method — статистический метод «Монте-Карло»
Multidimensional Scaling — многомерное шкалирование
Multiple — множественный
Multiple comparisons — множественные сравнения (средних)
Multivariate — многомерный
Multivariate Approach — многомерный подход
Nearest Neighbor — ближайшего соседа (метод кластеризации)
Negative — отрицательный
Next — следующий
Nominal — номинальная (шкала)
Nonparametric Test — непараметрический критерий
Normal Curve — нормальная кривая
Number — количество (численность), номер
Observed Frequency — наблюдаемая (эмпирическая) частота
Observed Prop. — наблюдаемое (эмпирическое) соотношение
One-Sample Kolmogorov-Smirnov Test — критерий (тест) Колмогорова-Смирнова
One-Sample T Test — критерий *t*-Стьюдента для одной выборки
One-tailed (one-sided) — односторонний критерий (для проверки односторонних гипотез); one-tailed — дословно однохвостый
One-Way ANOVA — однофакторный дисперсионный анализ
Ordinal (Rank order) — порядковая (ранговая) (шкала)
Paired-Samples T Test — критерий *t*-Стьюдента для зависимых выборок
Pairwise — попарно, попарный
Partial Correlation — частная корреляция (коэффициент)
Pearson Chi-Square — критерий хи–квадрат Пирсона
Pearson correlation — корреляция Пирсона
Pearson r — корреляция Писрона
Percentage — процент
Percentiles — процентили
Phi — фи–коэффициент сопряженности

Phi 4-points correlation — (четырехклеточный) фи—коэффициент сопряженности (Пирсона)
Pillai's Trace — след Пиллая
P-level — статистическая значимость (*р*-уровень)
Plot — диаграмма
Polynomial — полином (многочлен)
Post Hoc — апостериорный (после подтверждения гипотезы)
P-P Plot — график накопленных частот
Predicted (value, group) — предсказанное (значение, группа)
Principal Axis Factoring — факторный анализ методом главных осей
Principal Components — главных компонент (анализ)
Prior probabilities — априорные вероятности
Probability — вероятность
Probability of group membership — вероятность принадлежности к группе
Proximity — близость
PROXSCAL — программа неметрического многомерного шкалирования
Q-Q Plot — квантильный график
Quantile — квантиль
Quartile — квартиль
R Square (R2) — коэффициент детерминации (квадрат коэффициента корреляции)
Random (Factor) — случайный (фактор)
Range — размах
Rank — ранг
Rank (Cases) — ранжировать (объекты)
Ratio — абсолютная шкала (равных отношений)
Raw — строка
Raw Data — данные в строках (типа «объект-признак»)
Rectangular (matrix) — прямоугольная (матрица)
Regression — регрессионный
Regression Coefficient — коэффициент регрессии
Regression Line — линия регрессии
Related Samples — зависимые (связанные) выборки
Removal — удаление
Repeated Measures — повторные измерения
Repeated Measures ANOVA — дисперсионный анализ с повторными измерениями
Residual — остаток, отклонение (ошибка модели)
Residual analysis — анализ остатков (ошибок)
Rotation (Factors) — вращение (факторов)
Rotation Method — метод вращения
Row — строка
RSQ (R-square) — квадрат коэффициента корреляции, коэффициент дискриминации (r-квадрат)
Run — серия
Runs Test — критерий серий
Sample — выборка
Scaling model — модель шкалирования
Scatter Plot — диаграмма рассеивания
Scheffé test — метод Шеффе (множественных сравнений средних)
Scree plot — график собственных значений
Scree-test — критерий отсеивания
Scrollsheet — таблица результатов статистического анализа
Separate (Line) — отдельная (линия)
Set — множество
Shape — уточнить
Sig. — уровень значимости (р-уровень)
Sign — знак
Sign Test — критерий знаков
Significance Level — уровень статистической значимости (р-уровень)
Significant — статистически значимый
Similarity — сходство
Simple — простой
Simple Matching — простой коэффициент совстречаемости (бинарный)
Single Linkage — одиночной связи (метод кластеризации)
Size difference — величина различий (бинарная)
Skewness — асимметрия
Slope — уклон, наклон (прямой по отношению к оси X), значение коэффициента регрессии
Solution — решение (выбор)
Source — источник (изменчивости, влияния)
Spearman's rho (r) — корреляция Спирмена (коэффициент)
Specify — определять
Spreadsheet — электронная таблица для исходных данных
SPSS — компьютерная программа «Статистический пакет для социальных наук» (сокр.)
Square asymmetric (symmetric) matrix — квадратная асимметричная (симметричная) матрица
Squared (Euclidean distance) — квадрат (Евклидова расстояния)
S-stress convergence — величина сходимости s-стресса
Stage — ступень (этап, шаг)
Standard deviation) — стандартное отклонение (сигма)
Standard error — стандартная ошибка

Standardized (Beta-) Coefficients — стандартизированные (бета-) коэффициенты регрессии
STATISTICA — компьютерная программа для статистического анализа данных
Statistical Test — статистический критерий (тест)
Std. Deviation — стандартное отклонение (сигма)
Std. Error — стандартная ошибка
Step — шаг
Stepwise Method — пошаговый метод
Stimulus — стимулы
Stress — стресс, в многомерном шкалировании
Stub-and-Banner Tables — таблицы сопряженности для двух и более признаков
Subject — субъект
Subject Weight — индивидуальный вес
Subset — подмножество
Sum — сумма
Sum of Squares (SS) — сумма квадратов
Summary — итог
Suppress — скрыть
Symmetric — симметричная
Syntax — командный файл в SPSS
Table — таблица
Test — статистический критерий
Test Prop. — ожидаемое (теоретическое) соотношение
Test Statistics — результаты статистической проверки
Test Value — заданное (тестовое) значение
Tied Ranks — связанные (повторяющиеся) ранги
Ties (of ranks) — связи, повторы (рангов)
Total — итог (сумма)
Transform (Values, Measures) — преобразование (значений, мер)
Tree Diagram — древовидная диаграмма
Two-tailed (two-sided) — двусторонний критерий (для проверки двусторонних гипотез)
Univariate (procedure, approach) — одномерный (метод, подход)
Unrotated — до вращения
Unstandardized Coefficient — коэффициент регрессии (не стандартизированный)
Untie tied observation — разъединять (различать) связанные (одинаковые) данные
Unweighted least squares — метод не взвешенных наименьших квадратов
Valid n — число случаев, по которому были проведены расчеты
Value — значение
Var — переменная (сокр.)
Variable — переменная (признак)
Variable List — список переменный
Variance — дисперсия
Varimax — метод варимакс (вращения факторов)
Vertical — вертикальная (ориентация)
Weight — вес
Wilcoxon Signed-rank (Matched pairs) test — критерий T-Вилкоксона
Wilcoxon test — критерий Вилкоксона
Wilks' Lambda — лямбда Вилкса
Within — внутри
Within-Group — внутригрупповой
Within-Subject — внутри объектов (внутригрупповой)
XI — хи-квадрат критерий
Yates' corrected — с поправкой Йетса
Yule's Q — Юла *Q*-коэффициент
Z-score — *z*-значение

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

Л

М

Н

О

П

Р

С

Т

У

Ф

Х

Ц

Ч

Ш

Э

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Математические методы

Айвазян С. А. и др. Прикладная статистика. Классификация и снижение размерности. М., 1989.

Анастази А. Психологическое тестирование: В 2 кн. Кн. 1. М., 1982.

Боровиков В. Statistica. Искусство анализа данных на компьютере: Для профессионалов. СПб., 2003.

Бурлачук Л. Ф., Морозов С. М. Словарь-справочник по психологической диагностике. СПб., 1999.

Гайда В. К., Захаров В. П. Психологическое тестирование: Учеб. пособие. Л., 1982.

Гласс Дж., Стенли Дж. Статистические методы в педагогике и психологии. М., 1976.

Гмурман В. Е. Теория вероятностей и математическая статистика. М., 1997.

Гублер Е. В., Генкин А. А. Применение непараметрических критериев статистики в медико-биологических исследованиях. Л., 1973.

Гусев А. Дисперсионный анализ в экспериментальной психологии: Учебное пособие для студентов факультетов психологии... М., 2000.

Гусев А. Н., Измайлов Ч. А., Михалевская М. Б. Измерение в психологии: общий психологический практикум. М., 1997.

Дружинин В. Н. Психодиагностика общих способностей. М., 1996.

Дэйвисон М. Многомерное шкалирование: методы наглядного представления данных. М., 1988.

Дюк В. А. Компьютерная психодиагностика. СПб., 1994.

Закс Л. Статистическое оценивание. М., 1976.

Иберла К. Факторный анализ. М., 1980.

Ивантер Э. В., Коросов А. В. Основы биометрии. Петрозаводск, 1992.

Кендалл М., Стьюарт А. Статистические выводы и связи. М., 1973.

Кричевец А. Н. Математика для психологов: Учебник / А. Н. Кричевец, Е. В. Шикин, А. Г. Дьячков / Под ред. А. Н. Кричевца. М., 2003.

Куликов Л. В. Психологическое исследование. СПб., 1995.

Лаак Я. тер Психодиагностика: проблемы содержания и методов. М., Воронеж, 1996.

Мельников В. М., Ямпольский Л. Т. Введение в экспериментальную психологию личности. М., 1985.

Митина О. В., Михайловская И. Б. Факторный анализ для психологов. М., 2001.

Общая психодиагностика / Под ред. А. А. Бодалева, В. В. Столина. М., 1987.
Паповян С. С. Математические методы в социальной психологии. М., 1983.
Рунион Р. Справочник по непараметрической статистике. М., 1982.
Сидоренко Е. В. Методы математической обработки в психологии. СПб., 1996.
Справочник по прикладной статистике. В 2-х т. / Под ред. Э. Ллойда, У. Ледермана, Ю. Тюрина. М., 1989.
Суходольский Г. В. Математическая психология. СПб., 1997.
Суходольский Г. В. Основы математической статистики для психологов. СПб., 1998.
Суходольский Г. В. Математические методы психологии. СПб., 2003.
Тарасов С. Г. Основы применения математических методов в психологии. СПб., 1998.
Терехина А. Ю. Анализ данных методами многомерного шкалирования. М., 1986.
Факторный, дискриминантный и кластерный анализ / Дж.-О. Ким, Ч. У. Мьюллер, У. Р. Клекка и др. М., 1989.
Харман Г. Современный факторный анализ. М., 1972.
Шмелев А. Г. Психодиагностика личностных черт. СПб., 2002.

Организация и планирование исследования

Кэмпбелл Д. Модели экспериментов в социальной психологи и прикладных иследованях. М., 1980.
Мартин Д. Психологические эксперименты. СПб., 2002.
Солсо Р. Л., МакЛин М. К. Экспериментальная психология. СПб., 2003.

Андрей Дмитриевич Наследов

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ПСИХОЛОГИЧЕСКОГО ИССЛЕДОВАНИЯ

Анализ и интерпретация данных

Учебное пособие

Главный редактор *И. Авидон*
Заведующая редакцией *Т. Тулупьева*
Технический редактор *О. Колесниченко*
Художественный редактор *П. Борозенец*
Директор *Л. Янковский*

Подписано в печать 19.07.2004.
Формат $70\times100^1/_{16}$. Усл. печ. л. 31,6. Тираж 5000 экз. Заказ № 63.

Интернет-магазин: www.internatura.ru

ООО Издательство «Речь».
199178, Санкт-Петербург, ул. Шевченко, д. 3 (лит. "М"), пом. 1,
тел. (812) 323-76-70, 323-90-63, info@rech.spb.ru, www.rech.spb.ru
Представительство в Москве:
тел.: (095) 502-67-07, rech@online.ru

Отпечатано с готовых диапозитивов в ФГУП «Печатный двор»
Министерства РФ по делам печати, телерадиовещания
и средств массовых коммуникаций.
197110, Санкт-Петербург, Чкаловский пр., 15.



Наследов Андрей Дмитриевич — кандидат психологических наук, доцент факультета психологии СПбГУ. Область научных и педагогических интересов: организация и методы психологического исследования, современные методы анализа данных, история и методология психологии. На факультете психологии СПбГУ преподает с 1990 года; читает курсы: «Математические методы в психологии», «История психологии», «Статистические методы и математические модели в психологии», «Дизайн психологического исследования».

- Элементарные основы и первичные статистики.
- Проверка гипотез: методы статистического вывода.
- Многомерные методы и модели.
- Обработка данных на компьютере.

Стиль книги выбран с учетом того, что математические методы обычно вызывают большие трудности при изучении. Необходимые для понимания математические основы даются скорее на неформальном уровне — без детального изложения математического обоснования и выводов формул. Введение математических терминов сопровождается простыми примерами, а теоретические и математические объяснения даются на элементарном уровне.

А. Д. Наследов


ISBN 5-9268-0275-X
9785926802754